



Naafs, B. D. A., Gallego-Sala, A. V., Inglis, G. N., & Pancost, R. D. (2017). Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. *Organic Geochemistry*, 106, 48-56.
<https://doi.org/10.1016/j.orggeochem.2017.01.009>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.orggeochem.2017.01.009](https://doi.org/10.1016/j.orggeochem.2017.01.009)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.orggeochem.2017.01.009> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration

B.D.A. Naafs^{1*}, A.V. Gallego-Sala², G.N. Inglis¹, and R.D. Pancost¹

¹Organic Geochemistry Unit, School of Chemistry and Cabot Institute, University of Bristol, Bristol, UK

²Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

*Corresponding author. E-mail address: david.naafs@bristol.ac.uk

Abstract

Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are increasingly used to reconstruct past terrestrial temperature and soil pH. Here we compare all available modern soil brGDGT data (n=350) to a wide range of environmental parameters to obtain new global temperature calibrations.

We show that soil moisture index (MI), a modeled parameter that also takes potential evapotranspiration into account, is correlated to the 6-methyl brGDGT distribution but does not significantly control the distribution of 5-methyl brGDGTs. Instead, temperature remains the primary control on 5-methyl brGDGTs. We propose the following global calibrations: $MAAT_{soil} = 40.01 \times MBT'_{5me} - 15.25$ (n=350, $R^2 = 0.60$, RMSE = 5.3 °C) and growing degree days above freezing ($GDD_{0\ soil} = 14344.3 \times MBT'_{5me} - 4997.5$ (n=350, $R^2 = 0.63$, RMSE = 1779 °C).

Recent studies have suggested that factors other than temperature can impact arid and/or alkaline soils dominated by 6-methyl brGDGTs. As such, we develop new global temperature calibrations using samples dominated by 5-methyl brGDGTs only ($IR_{6me} < 0.5$). These new calibrations have significantly improved correlation coefficients and lower root mean square errors (RMSE) compared to the global calibrations: $MAAT_{soil}' = 39.09 \times MBT'_{5me} - 14.50$ (n=177, $R^2 = 0.76$, RMSE = 4.1 °C) and $GDD_{0\ soil}' = 13498.8 \times MBT'_{5me} - 4444.5$ (n=177, $R^2 = 0.78$, RMSE = 1326). We suggest that these new calibrations should be used to reconstruct terrestrial

climate in the geological past; however, care should be taken when employing these calibrations outside the modern calibration range.

1. Introduction

Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are membrane-spanning lipids produced by bacteria, presumably acidobacteria (Weijers et al., 2009; Sinninghe Damsté et al., 2011). First discovered in a Dutch peat (Sinninghe Damsté et al., 2000), brGDGTs are ubiquitous in mesophilic settings (Schouten et al., 2013) such as soils (Weijers et al., 2006b; Peterse et al., 2012), lakes (Pearson et al., 2011; Schoon et al., 2013; Li et al., 2016), rivers (De Jonge et al., 2014b), marine sediments (Hopmans et al., 2004; Fietz et al., 2012), and peat deposits (Weijers et al., 2006a; Huguet et al., 2010; Zheng et al., 2015; Naafs et al., 2017). Recent advances in analytical methods (De Jonge et al., 2013; Yang et al., 2015; Hopmans et al., 2016) have revealed the existence of a wide range of brGDGTs in mineral soils (Fig. S1), varying in the number of methyl branches (between 4 and 6), the carbon position of methyl branches (C5 and C6 position), and number of cyclopentane moieties (between 0 and 2).

Over the last decade brGDGTs have become of great interest to the organic geochemistry and paleoclimate communities because their distribution (degree of cyclisation and methylation) correlates with soil-pH and mean annual air temperature. This was originally expressed by Weijers et al. (2007b) in the MBT/CBT-proxy in a global soil dataset and redefined by Peterse et al. (2012) with the MBT'/CBT-proxy:

$$(1) MBT' = \frac{Ia + Ib + Ic}{Ia + Ib + Ic + IIa + IIa' + IIb + IIb' + IIc + IIc' + IIIa + IIIa'}$$

$$(2) CBT = -\log\left(\frac{Ib + IIb + IIb'}{Ia + IIa + IIa'}\right)$$

$$(3) MAT (^{\circ}C) = 0.81 - 5.67 \times CBT + 31.0 \times MBT' \quad (n = 176, R^2 = 0.59, RSME = 5.0 ^{\circ}C)$$

Consequently, the brGDGT-based MBT(')/CBT-proxy has been increasingly applied to proximal marine and lake sediments as well as loess and paleosols to gain insights into past terrestrial temperatures (Weijers et al., 2007a; Pancost et al., 2013; Schouten et al., 2013; Peterse et al., 2014; Lu et al., 2016).

De Jonge et al. (2013; 2014a) recently demonstrated that 5-methyl penta- and hexamethylated brGDGTs, used to calculate the CBT and MBT(') indices, co-elute with newly identified 6-methyl brGDGTs. Re-evaluation of the global soil calibration

dataset in this context removed the pH dependence upon the degree of methylation of brGDGTs (De Jonge et al., 2014a) and suggested that the abundance of 6-methyl brGDGTs is influenced predominantly by pH (De Jonge et al., 2014a; Xiao et al., 2015). Excluding the 6-methyl brGDGTs from the regressions, De Jonge et al. (2014a) developed two new types of equations with a dependence of 5-methyl brGDGTs on temperature alone; one based on the degree of methylation of 5-methyl branched tetraethers (MBT'_{5ME}):

$$(4) MBT'_{5ME} = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa + IIb + IIc + IIIa)}$$

$$(5) MAT = -8.57 + 31.45 \times MBT'_{5ME} \quad (n = 231, R^2 = 0.64, RMSE = 4.9 \text{ } ^\circ C)$$

And another based on a multiple linear regression using the relative abundance of specific 5-methyl brGDGTs (MAT_{mr}).

$$(6) MAT_{mr} = 7.17 + 17.1 \times \{Ia\} + 25.9 \times \{Ib\} + 34.4 \times \{Ic\} - 28.6 \times \{IIa\}$$

$$(n = 231, R^2 = 0.67, RMSE = 4.7 \text{ } ^\circ C)$$

(Note that the calibration statistics (n , R^2 , and RMSE) given in De Jonge et al. (2014a) were recently corrected (De Jonge et al., 2016)).

Although the latest calibrations improved the correlation coefficient and root mean square error (RMSE) of brGDGT temperature and pH proxies, the relatively large scatter in the global calibration indicates the potential influence of additional environmental parameters, such as precipitation and soil moisture content (SMC) on brGDGT distributions (Weijers et al., 2011; Anderson et al., 2014; Ding et al., 2015; Xiao et al., 2015; Yang et al., 2015; Dang et al., 2016; Lei et al., 2016). For example, Dang et al. (2016) recently showed that the soil brGDGT distribution changed along a short transect with varying soil moisture content.

Although previous studies have argued that mean annual temperature (MAT), soil pH, and precipitation (MAP) are the key environmental parameters influencing brGDGT distributions in the global soil database, the influences of other parameters such as growing degree days (GDD) have not been considered in the context of the global dataset. Further more, although some studies have investigated the impact of soil moisture on the brGDGT distribution at a regional level (Dirghangi et al., 2013; Dang et al., 2016), this has not been studied in a global context. This is important as both soil moisture and GDD are potentially better indicators of the growth temperature and moisture content experienced by brGDGT-producing bacteria living in soils (McMaster and Wilhelm, 1997; Gallego-Sala et al., 2010). An additional

source of error can derive from the instrumental temperature data used in calibrations, because this comprises a mix of global databases and local weather station data and potentially there is an offset between air temperature and soil temperature (Weijers et al., 2007b; Peterse et al., 2012). In addition, the existing mineral soil calibrations indicate a bias in the coldest soils with brGDGT-based temperatures up to 15 °C higher than observed mean annual air temperature (De Jonge et al., 2014a).

Here we compile and revisit all available soil brGDGTs data from around the world and calibrate these against a range of environmental parameters obtained from a simple bioclimatic model (PeatStash) that calculates bioclimatic variables using long-term mean monthly values of temperature, precipitation and the fraction of possible sunshine hours (Gallego-Sala et al., 2010). We use this to assess the environmental controls on brGDGTs in soils and re-define the global soil temperature proxies. Although several studies advocate the use of local calibrations (e.g., Ding et al., 2015; Yang et al., 2015), the earth's climate system was significantly different during the geological past (e.g., during the Eocene). As such, the application of local calibrations should remain limited to recent (i.e. Quaternary) sediments when environmental conditions were likely similar to those covered in the local calibration dataset. For deep time application, global calibrations are required as these incorporate all modern-day climate zones.

2. Material and methods

2.1. Material

We use the distribution of brGDGTs in the global soil dataset compiled by De Jonge et al., (2014a), based on a sample set generated previously (Weijers et al., 2007b; Peterse et al., 2012). This is supplemented with data from Chinese soils (Ding et al., 2015; Xiao et al., 2015; Yang et al., 2015; Lei et al., 2016). Although other data sets containing modern brGDGT distributions in soils exist, these do not separate the 5- and 6-methyl brGDGTs and were therefore not included. The revised dataset from De Jonge et al., (2014a) consists of 239 samples from across the globe. We exclude 13 samples either because the altitude of the soil sample was unknown (Peterse, personal communication December 2015) or because the altitude in PeatStash for that location was significantly different compared to that of the soil sample and no altitude correction could be made. Combined with 27 samples from the Qinghai–Tibetan Plateau (Ding et al., 2015), 27 samples from across the 400 mm isoline of mean

annual precipitation in China (Xiao et al., 2015), 26 samples from Mt. Shennongjia in China (Yang et al., 2015), and 44 samples from the Henan and Yunnan provinces in China (Lei et al., 2016), we use a total of 350 soil samples (Fig. 1). The global dataset consists of data measured in different laboratories. Although there are no data available for the interlaboratory variation of soil brGDGT-based indices, compiling data from different laboratories could introduce additional variation.

2.2. Environmental parameters

We used bioclimatic variables obtained from a simple bioclimatic model (PeatStash). PeatStash calculates these variables globally with a 0.5 degree spatial resolution (Gallego-Sala and Prentice, 2013). The calculations are based on long-term mean climatology data, obtained by interpolating long-term mean weather station climatology (temperature, precipitation and the fraction of possible sunshine hours) from around the world for the period 1931-1960 (Climate 2.2 available online <http://www.pik-potsdam.de/~cramer/climate.html>). If the altitude of the grid cell was significantly different (> 250 meter) from that reported for a soil sample, a nearby grid cell with an altitude difference < 250 m was used for comparison. For soil altitude transects (e.g., Peterse et al., 2009), we generate temperature transects using PeatStash, but these could not be used to calculate precipitation or moisture index across transects. The temperature dataset thus consists of 350 soils, whereas the dataset for MAP and moisture index consists of 275 soils.

brGDGT distributions were compared to the following climatological data, obtained using PeatStash (Gallego-Sala et al., 2010; Gallego-Sala and Prentice, 2013): mean annual air temperature (MAAT), mean warmest month temperature (MWMT), mean annual precipitation (MAP), moisture index (MI), and growing degree days above 0°C (GDD_0). The MI is defined as annual precipitation over annual potential evapotranspiration (P/PET) (Gallego-Sala et al., 2010). MI provides a better measure of water availability compared to MAP as it takes into account the large difference in evaporative demand between different climate regimes. Values < 1 are indicative of dry soils whereas values > 3 are encountered in the wettest areas on earth. GDD_0 is defined as the yearly cumulated daily average temperature of the daily maximum and minimum temperature for average temperatures $> 0^{\circ}\text{C}$. GDD_0 is a measure of annual soil heat accumulation and widely used to predict the timing of biological processes (Kaplan et al., 2003). A high value is indicative of a (sub)tropical climate and a low

value for polar/tundra climates. At MAAT > 15 °C, GGD₀ is linearly correlated with MAAT following 365 (the number of days in a year) multiplied by MAAT (the mean annual air temperature at a given location).

2.3. Statistical methods

Instead of simple linear regression, we use Deming regressions. The advantage of Deming regressions is that they account for error in both x and y-axis, meaning both the proxy (e.g. MBT_{5me}') and environmental parameter (e.g., MAAT) (Adcock, 1878). For this purpose we used RStudio (RStudio Team, 2015) and the Method Comparison Regression (MCR) package (Manuilova et al., 2014), which are freely available to download¹. The Rscript and data are available in the supplements for future users. The errors associated with the proxy measurements (e.g. MBT_{5me}') and environmental parameters (e.g., MAAT) are independent and assumed to be normally distributed. In order to calculate the ratio of their variances (δ), needed to calculate a Deming regression, we assumed that the standard deviation (σ) of the environmental data MAAT, GDD₀, and pH are 1.5 °C, 547.5 °C (365 x 1.5 °C) and 0.25, respectively. For the brGDGT-based proxies (e.g. MBT_{5me}') we assumed a σ of 0.05. This results in a δ of 0.0011 for the MBT_{5me}'/MAAT calibration, 8.3 x 10⁻⁷ for the MBT_{5me}'/GDD₀ calibration, and 0.04 for the pH calibrations (see supplementary information), respectively. Residuals were calculated using

$$(7) \text{Residual}_y = y_{\text{observed}} - y_{\text{predicted}}$$

The root mean square error (RMSE) for y, the predictive error for the environmental parameter of interests (e.g., MAAT), was calculated using

$$(8) \text{RSME}_y = \sqrt{\frac{\sum_{x=1}^n (y_{x,\text{observed}} - y_{x,\text{predicted}})^2}{n}} \times \frac{n}{df}$$

Where *df* stands for degrees of freedom, which in this case is n-1.

3. Results and Discussion

Previous studies suggested that the distribution of brGDGTs in soils is controlled predominantly by MAAT and soil pH (Weijers et al., 2007b; Peterse et al., 2012; De Jonge et al., 2014a). Other studies have explicitly focused on the pH dependence of brGDGTs in the global data set, showing that the relative abundance of 6-methyl

¹ <https://www.rstudio.com> and <https://cran.r-project.org/web/packages/mcr/index.html>

brGDGTs is positively correlated to pH (Ding et al., 2015; Xiao et al., 2015). Guided by these results we focus on the influence of a range of environmental parameters on the fractional abundance of brGDGTs.

3.1 brGDGTs versus temperature

When the fractional abundances of brGDGTs from all 350 samples are plotted versus mean annual air temperature (MAAT), it is clear that only 5-methyl brGDGTs lacking cyclopentane moieties (i.e. brGDGT-Ia, -IIa, and -IIIa) are significantly ($R^2 > 0.2$, $p < 0.001$) correlated to MAAT (Fig. 2). brGDGT-Ia is positively correlated with MAAT ($R^2 = 0.38$, $p < 0.001$), whereas brGDGT-IIa ($R^2 = 0.20$, $p < 0.001$) and -IIIa ($R^2 = 0.36$, $p < 0.001$) are negatively correlated with MAAT. These values (R^2 of 0.38, 0.20 and 0.36) are similar to those reported by De Jonge et al. (2014a) for these compounds with R^2 of 0.34, 0.43, and 0.43, respectively, although the dataset used here is larger. The correlation between brGDGT-IIa and MAAT is lower in the dataset used here. Together these results confirm the fundamental dependence of brGDGT distributions on temperature with lower temperatures associated with a higher degree of methylation. This was originally proposed by Weijers et al. (2007b) who argued that additional methyl groups result in a more loose packing of brGDGTs, allowing bacteria to maintain membrane fluidity at lower temperature, similar to what is seen in fatty acids synthesized by bacteria (Russell, 1984). 6-methyl brGDGTs and brGDGTs containing cyclopentane moieties are not significantly correlated to MAAT ($R^2 < 0.11$).

3.2 brGDGTs versus precipitation and soil moisture index

As observed by Weijers et al. (2007b), the fractional abundances of several brGDGTs are also significantly correlated to mean annual precipitation (MAP). In our data set ($n = 275$) the highest correlations are found for the relative abundance of 5-methyl brGDGT-Ia ($R^2 = 0.48$, $p < 0.001$) and -IIIa ($R^2 = 0.21$, $p < 0.001$), as well as 6-methyl brGDGT-IIa' ($R^2 = 0.35$, $p < 0.001$) and -IIIa' ($R^2 = 0.24$, $p < 0.001$) (Fig. 2). Intriguingly, the correlation of brGDGT-Ia with MAP ($R^2 = 0.48$) is higher than that with MAAT ($R^2 = 0.38$).

Crucially, MAP does not reflect the soil moisture content experienced by soil bacteria as the latter also depends on evaporation and transpiration. To explore this further, we compared the brGDGTs distribution to the soil moisture index calculated

by PeatStash (Fig. 2), although we want to stress that this approach does not take into local variations (e.g. microtopography, etc) that might be important in determining soil moisture content for a given sample. A soil moisture index of < 1 indicates that the potential annual evapotranspiration (the combined effect of evaporation and transpiration) is higher than annual precipitation and hence a dry soil, whereas a value > 1 indicates the opposite and suggests a wet soil. The correlation of brGDGT-Ia with soil moisture index is significantly lower ($R^2 = 0.26$, $p < 0.001$) than that observed for MAP ($R^2 = 0.48$, $p < 0.001$), and the correlation for brGDGT-IIIa drops below 0.2. Therefore, we suggest that the high correlation between MAP and 5-methyl brGDGTs-Ia is partly due to the correlation between MAP and MAAT ($R^2 = 0.51$ for this dataset). In contrast, for 6-methyl brGDGTs (IIa' and IIIa'), the correlation with MI is 0.29 and 0.20, respectively ($p < 0.001$), and similar to that of MAP. Taken together this suggests a control of moisture content on the abundance of 6-methyl brGDGTs in soils. These results are supported by a recent study from Dang et al. (2016) who demonstrated that the brGDGT distribution in Chinese soils depends on soil moisture content with a higher amount of brGDGT-IIa' and -IIIa' in dry mineral soils compared to wet mineral soils.

3.3 brGDGTs versus warmest mean month temperature

Seasonality is not generally considered to impact the brGDGT distribution in soils, but there is only limited data to support this. A study of several mid-latitude soils argued against seasonal changes in brGDGT production as: 1) MBT/CBT-derived temperatures yield similar temperature estimates throughout the year, and 2) the concentration of brGDGTs remained constant through the year, indicating a slow turnover time of brGDGT-producing bacteria on the order of ~ 20 years (Weijers et al., 2011). However, these results do not preclude a systematic bias in brGDGT production (and therefore environmental influence) towards a particular season in high-latitude regions, most likely the warm season as bacterial growth is greater at higher temperature. Indeed, much higher turnover rates (< 2 years) and a bias in brGDGT distribution towards the warmest month of the year were recently inferred in a French peatland (Huguet et al., 2013).

To examine the influence of warm season temperature on GDGT distributions we compared the brGDGT distribution to mean warmest month temperature (MWMT). The aim was to investigate whether there is a better correlation with

MWMT compared to that observed for MAAT (Fig. 3). However, the correlation coefficients for MWMT (e.g. R^2 for brGDGT-Ia is 0.22) are overall lower than those for MAAT (R^2 for brGDGT-Ia is 0.38), which would suggest that on a global basis there is no bias towards the warm season in brGDGT distribution. The same results are obtained when only samples from soils with MAAT < 5 °C are used (Fig. 3). This is further supported by a recent study that reported no difference in brGDGT distribution between Chinese soils with contrasting seasonality (Lei et al., 2016).

3.4 brGDGTs versus growing degree days above freezing

Although our results do not indicate a global bias in seasonality, the season of brGDGT production could still be dependent upon latitude. For example, bacteria in a tropical and temperate soil are likely to grow throughout the year with temperatures always above freezing, whereas those in a high latitude soil could be heavily biased to those months when soil temperatures are above zero. Indeed, it is hard to envision that in the high-latitudes, where winters are characterized by subzero temperatures, that brGDGTs are produced in equal amounts throughout the year. To explore this further, we compared the brGDGTs distributions to growing degree days above freezing (GDD_0), a measure of the cumulative temperature (in °C) above zero a soil experiences over the year.

MAAT has previously been considered to reflect the temperature that soil bacteria experience. However, this does not reflect the time and intensity at which a given soil remains above freezing. In temperate and polar climates with MAAT < 15 °C, GDD_0 is more indicative of the cumulative heat a soil experiences. For example, soils from central Kazakhstan and Newfoundland (at 49 °N latitude) both experience a MAAT of ~3.5 °C. However, Kazakhstan is characterized by a continental climate with extremely cold winters and hot summers, whereas Newfoundland has a maritime climate with much less extreme seasonal variation. GDD_0 differentiates the two climates as the warm summers in Kazakhstan lead to a GDD_0 for this region of around 2890 °C (cumulative degrees centigrade above zero over one year), much higher than the value of 1930 °C for Newfoundland.

As with MAAT, when GDD_0 is compared to the brGDGT distribution only 5-methyl brGDGTs lacking cyclopentane moieties (i.e. brGDGT-Ia, -IIa, - and -IIa) are significantly ($R^2 > 0.2$) correlated (Fig. 3). brGDGT-Ia is positively correlated with GDD_0 ($R^2 = 0.42$, $p < 0.001$), whereas brGDGT-IIa ($R^2 = 0.24$, $p < 0.001$) and -IIIa

($R^2 = 0.38$, $p < 0.001$) are negatively correlated. These R^2 values are slightly higher than those found for MAAT and higher than those found for MWMT (Fig. 2).

3.5 Redefining temperature calibrations using MBT'_{5me} and Deming regressions

Our results confirm that the degree of methylation is significantly correlated with temperature, either using MAAT or GDD_0 . Following previous studies (De Jonge et al., 2014a; Ding et al., 2015) we calculated the modified methylation index of 5-methyl branched tetraethers MBT'_{5me} (see equation 4) and calibrate this against MAAT and GDD_0 using Deming regressions (Fig. 4a and 4d).

This results in the following two Deming temperature regressions:

$$(9) MAAT_{soil} (^{\circ}C) = 40.01 \times MBT'_{5me} - 15.25 \quad (n = 350, R^2 = 0.60, RMSE = 5.3 ^{\circ}C)$$

$$(10) GDD_{0\ soil} = 14344.3 \times MBT'_{5me} - 4997.5 \quad (n = 350, R^2 = 0.63, RMSE = 1779 ^{\circ}C)$$

Overall the GDD_0 calibration performs slightly better than the MAAT calibration as it has a slightly higher R^2 . Although the slope and intercept of $MAAT_{soil}$ are different, the R^2 and RMSE are similar compared to those reported for the MBT_{5me} -MAT calibration by De Jonge et al. (2014a) ($n = 231$, $R^2 = 0.64$, and $RMSE = 4.9 ^{\circ}C$), but lower compared to those reported by Ding et al. (2015) ($n = 249$, $R^2 = 0.70$, and $RMSE = 4.7 ^{\circ}C$). Nonetheless, there is still significant scatter in our revised calibrations (Fig. 4b and 4e). Interestingly, the calibrations versus MAAT are characterized by relatively large residuals at lower temperatures (Fig. 4b). Specifically, the brGDGT distribution (MBT_{5me}') overestimates MAAT at lower temperatures and may be related to a seasonal production bias in high-latitudes sites. Indeed these low temperature residuals are reduced when GDD_0 is applied instead of MAAT (Fig. 4e).

An additional problem with both calibrations is that MBT'_{5me} reaches 1 at a MAAT of $24.8 ^{\circ}C$ and GDD_0 at $9347 ^{\circ}C$, thereby compromising the application of these calibrations to (sub)tropical settings both in the recent past but especially in the geological past when terrestrial temperatures were generally higher (e.g., Huber and Caballero, 2011).

3.6 Temperature calibrations using multiple linear regressions

As MBT'_{5me} reaches 1 at a relatively low MAAT (22.9 °C) in the temperature calibration of De Jonge et al. (2014a), the authors suggested that a multiple linear regression (MLR)-based calibration, based upon the fractional abundances of brGDGTs-Ia, -Ib, -Ic, and -IIa (eq. 6), was a more suitable choice for paleoclimate studies. As such, we have also explored the performance of MLRs in our expanded mineral soil dataset. The optimal MLRs are

$$(11) MAAT_{mlr\ soil} (^{\circ}C) = 19.8 \times \{Ia\} + 31.1 \times \{Ib\} - 23.4 \times \{IIa\} + 4.32 \quad (n = 350, R^2 = 0.62, RMSE = 4.7^{\circ}C)$$

$$(12) GDD_{0\ mlr\ soil} = 6152.9 \times \{Ia\} + 8272.1 \times \{Ib\} - 8015.8 \times \{IIa\} + 2509.4 \quad (n = 350, R^2 = 0.68, RMSE = 1319)$$

Adding additional compounds inflates the p -level of the slopes and intercept to values > 0.01 . The advantage of using a MLR model is that the correlations (R^2) improve and RMSEs decrease compared to the MBT'_{5me} calibrations. However, the MLRs 1) reach saturation (100% brGDGT-Ia) around 24-25 °C, similar to the MBT'_{5me} calibration, 2) do not account for the error in both the proxy and environmental parameter as Deming regressions do, and 3) are characterized by structural residuals, which are the most significant at low MAAT (Fig. 4c and 4f). We therefore suggest the MBT'_{5me} calibrations (eq. 9 and 10) are the optimal global calibrations. However, the amount of scatter in the calibration remain large, indicating that additional parameters may influence the total brGDGT distributions.

3.7 Temperature calibration excluding samples dominated by 6-methyl brGDGTs

Numerous studies have shown that there is a poor correlation between the methylation of brGDGTs and MAAT in arid and/or alkaline soils (Peterse et al., 2012; Dirghangi et al., 2013; Anderson et al., 2014; Zell et al., 2014; Ding et al., 2015; Yang et al., 2015). The reason for the apparent control of moisture (or other related environmental parameters) on the abundance of 6-methyl brGDGTs is unknown. Soil moisture is correlated to pH with dry soils predominantly being alkaline. Given that the abundance of 6-methyl brGDGTs is highly correlated to pH (see Fig. 5 and supplementary information), this might explain the correlation. In fatty acids, the position of methyl groups impacts membrane fluidity, with anteiso chains (methyl on C2 position) inducing a greater degree of fluidity compared to iso chains (methyl on

C1 position) (Denich et al., 2003 and references therein). Potentially the same applies to brGDGTs, whereby shifting a methyl group from C5 to C6 leads to a greater membrane fluidity in arid and/or alkaline soils. An alternative explanation could be that different communities that thrive at different pH and soil moisture content produce a different distribution of 5- and 6-methyl brGDGTs.

Using a transect of varying soil moisture content (SMC) in Chinese soils, Dang et al. (2016) demonstrated that SMC has an impact on the distribution of brGDGTs, in particular 6-methyl brGDGTs. Using the ratio of 6- over 5-methyl brGDGTs (IR_{6me}) they proposed that MBT' is only significantly correlated to MAAT in mineral soils with $IR_{6me} \leq 0.5$.

$$(13) IR_{6me}$$

$$= \frac{\{IIa'\} + \{IIb'\} + \{IIc'\} + \{IIIa'\} + \{IIIb'\} + \{IIIc'\}}{\{IIa'\} + \{IIb'\} + \{IIc'\} + \{IIIa'\} + \{IIIb'\} + \{IIIc'\} + \{IIIc'\} + \{IIIc'\} + \{IIIc'\} + \{IIa\} + \{IIb\} + \{IIc\} +}$$

Based on this observation, we evaluated the influence of IR_{6me} on the correlation coefficient (R^2) between MBT_{5me}' and MAAT (Fig. 6). In the global soil dataset the correlation coefficient (R^2) between MBT_{5me}' and MAAT decreases significantly from 0.76 to 0.67 when the threshold for IR_{6me} increases from < 0.5 to < 0.6 , similar to observations from the Chinese soil transect. These results suggest that the temperature dependence of brGDGTs in soils with a high amount of 6-methyl brGDGTs over 5-methyl brGDGTs (mainly arid/alkaline soils) is different. We therefore excluded samples with $IR_{6me} > 0.5$. From the total of 350 soil samples; roughly half (177) have $IR_{6me} < 0.5$, and these are mostly from acidic soils (Fig. 5). This yields a significant improvement in the correlations between individual brGDGT-Ia, -IIa, and -IIIa and MAAT (Fig. 7; R^2 values of 0.67, 0.68, and 0.51, respectively) and leads to significantly improved calibrations with higher R^2 and lower RMSEs (Fig. 8). This applies to both Deming (Eq. 14 and 15) and multiple linear regressions (Eq. 16 and 17).

$$(14) MAAT_{soil 5me} (^{\circ}C) = 39.09 \times MBT'_{5me} - 14.50 \quad (n = 177, R^2 = 0.76, RMSE = 4.1 ^{\circ}C)$$

$$(15) GDD_{0 soil 5me} = 13498.8 \times MBT'_{5me} - 4444.5 \quad (n = 177, R^2 = 0.78, RMSE = 1326)$$

$$(16) MAAT_{mlr\ soil\ 5me} (^{\circ}C) = 14.7 \times \{Ia\} - 31.7 \times \{IIa\} + 10.0 \quad (n = 177, R^2 = 0.77, RMSE = 3.8 ^{\circ}C)$$

$$(17) GDD_{0\ mlr\ soil\ 5me} = 4881 \times \{Ia\} - 10112 \times \{IIa\} + 3942 \quad (n = 177, R^2 = 0.82, RMSE = 1079)$$

Adding additional compounds to the MLRs does not improve the correlations and inflates the p -level of the slopes and intercept to values > 0.01 . As explained previously, the MLRs do not take the error in both proxy and environmental parameter into account, only perform slightly better than the MBT'_{5me} calibrations, and the MLR calibration of of MAAT (eq. 16) is characterized by structural residuals, especially at the low temperature end (Fig. 8c). As such we suggest that the MBT'_{5me} calibrations (eq. 14 and 15) are the best choice for paleoclimate reconstructions. However, both set of calibrations continue to saturate at temperatures of around 24-25 °C, implying that their application to past greenhouse climates has to be undertaken with caution.

These improvements over earlier calibrations imply that sample sets need to be screened for the abundance of significant amounts of 6-methyl brGDGTs prior to MAAT determinations. This will be particular important for archives from (semi-) arid regions such as loess and paleosols. It is important to note that in paleoclimate archives such as marine sediments, brGDGTs might be derived from a mixture of sources. This means that although these archives overall might be characterized by $IR_{6me} < 0.5$, there could be a contribution of soils with $IR_{6me} > 0.5$.

We envision that the ability to calculate growing degree days will be of particular interest to climate modelers as GDD_0 is more indicative of the seasonal temperature cycle than MAAT, especially at the high latitudes. Information about past seasonal temperatures and summer intensity is not readily available and provides a clear advantage of our calibration over other temperature proxies (marine and terrestrial).

4. Conclusions

The distribution of brGDGTs in soils has been shown previously to depend on environmental parameters such as mean annual air temperature (MAAT) and pH, but significant scatter in the existing calibrations suggests additional controls. Combining all available data, here we compare the brGDGT distribution to a range of

environmental parameters obtained from a globally integrated data set. In agreement with previous studies, we demonstrate that the distribution of 5-methyl brGDGTs depends primarily on temperature. Excluding samples from arid and/or alkaline soils dominated by 6-methyl brGDGTs significantly improves the correlation with temperature and growing degree days above zero (GDD₀). Guided by these results we provide new temperature calibrations. These new regressions have significantly improved correlation coefficients and lower root mean square errors (RMSE) compared to the existing global calibrations. We suggest that these new calibrations should be used to reconstruct terrestrial climate during the geological past, but caution should be taken when applying these calibrations to past greenhouse periods.

Acknowledgements

This research was funded through the advanced ERC grant 'The greenhouse earth system' (T-GRES, project reference 340923). R.D.P. acknowledges the Royal Society Wolfson Research Merit Award.

References

- Adcock, R.J., 1878. A Problem in Least Squares. *The Analyst* **5**, 53-54.
- Anderson, V.J., Shanahan, T.M., Saylor, J.E., Horton, B.K., Mora, A.R., 2014. Sources of local and regional variability in the MBT' /CBT paleotemperature proxy: Insights from a modern elevation transect across the Eastern Cordillera of Colombia. *Organic Geochemistry* **69**, 42-51.
- Dang, X., Yang, H., Naafs, B.D.A., Pancost, R.D., Evershed, R.P., Xie, S., 2016. Direct evidence of moisture control on the methylation of branched glycerol dialkyl glycerol tetraethers in semi-arid and arid soils. *Geochimica et Cosmochimica Acta* **189**, 24-36.
- De Jonge, C., Hopmans, E.C., Stadnitskaia, A., Rijpstra, W.I.C., Hofland, R., Tegelaar, E., Sinninghe Damsté, J.S., 2013. Identification of novel penta- and hexamethylated branched glycerol dialkyl glycerol tetraethers in peat using HPLC-MS², GC-MS and GC-SMB-MS. *Organic Geochemistry* **54**, 78-82.
- De Jonge, C., Hopmans, E.C., Zell, C.I., Kim, J.-H., Schouten, S., Sinninghe Damsté, J.S., 2014a. Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: implications for palaeoclimate reconstruction. *Geochimica et Cosmochimica Acta* **141**, 97-112.
- De Jonge, C., Stadnitskaia, A., Hopmans, E.C., Cherkashov, G., Fedotov, A., Sinninghe Damsté, J.S., 2014b. In situ produced branched glycerol dialkyl glycerol

tetraethers in suspended particulate matter from the Yenisei River, Eastern Siberia.
Geochimica et Cosmochimica Acta **125**, 476-491.

De Jonge, C., Hopmans, E.C., Zell, C.I., Kim, J.-H., Schouten, S., Sinninghe Damsté, J.S., 2016. Corrigendum to “Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction” [Geochim. Cosmochim. Acta 141 (2014) 97–112]. *Geochimica et Cosmochimica Acta* **183**, 276-278.

Denich, T.J., Beaudette, L.A., Lee, H., Trevors, J.T., 2003. Effect of selected environmental and physico-chemical factors on bacterial cytoplasmic membranes. *Journal of Microbiological Methods* **52**, 149-182.

Ding, S., Xu, Y., Wang, Y., He, Y., Hou, J., Chen, L., He, J.S., 2015. Distribution of branched glycerol dialkyl glycerol tetraethers in surface soils of the Qinghai-Tibetan Plateau: implications of brGDGTs-based proxies in cold and dry regions. *Biogeosciences* **12**, 3141-3151.

Dirghangi, S.S., Pagani, M., Hren, M.T., Tipple, B.J., 2013. Distribution of glycerol dialkyl glycerol tetraethers in soils from two environmental transects in the USA. *Organic Geochemistry* **59**, 49-60.

Fietz, S., Huguet, C., Bendle, J., Escala, M., Gallacher, C., Herfort, L., Jamieson, R., et al., 2012. Co-variation of crenarchaeol and branched GDGTs in globally-distributed marine and freshwater sedimentary archives. *Global and Planetary Change* **92–93**, 275-285.

Gallego-Sala, A.V., Clark, J.M., House, J.I., Orr, H.G., Prentice, I.C., Smith, P., Farewell, T., et al., 2010. Bioclimatic envelope model of climate change impacts on blanket peatland distribution in Great Britain. *Climate Research* **45**, 151-162.

Gallego-Sala, A.V., Prentice, C.I., 2013. Blanket peat biome endangered by climate change. *Nature Climate Change* **3**, 152-155.

Hopmans, E.C., Weijers, J.W.H., Schefuß, E., Herfort, L., Sinninghe Damsté, J.S., Schouten, S., 2004. A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. *Earth and Planetary Science Letters* **224**, 107-116.

Hopmans, E.C., Schouten, S., Sinninghe Damsté, J.S., 2016. The effect of improved chromatography on GDGT-based palaeoproxies. *Organic Geochemistry* **93**, 1-6.

Huber, M., Caballero, R., 2011. The early Eocene equable climate problem revisited. *Climate of the Past* **7**, 603-633.

Huguet, A., Fosse, C., Laggoun-Défarge, F., Toussaint, M.-L., Derenne, S., 2010. Occurrence and distribution of glycerol dialkyl glycerol tetraethers in a French peat bog. *Organic Geochemistry* **41**, 559-572.

- Huguet, A., Fosse, C., Laggoun-Défarge, F., Delarue, F., Derenne, S., 2013. Effects of a short-term experimental microclimate warming on the abundance and distribution of branched GDGTs in a French peatland. *Geochimica et Cosmochimica Acta* **105**, 294-315.
- Kaplan, J.O., Bigelow, N.H., Prentice, I.C., Harrison, S.P., Bartlein, P.J., Christensen, T.R., Cramer, W., et al., 2003. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research: Atmospheres* **108**.
- Lei, Y., Yang, H., Dang, X., Zhao, S., Xie, S., 2016. Absence of a significant bias towards summer temperature in branched tetraether-based paleothermometer at two soil sites with contrasting temperature seasonality. *Organic Geochemistry* **94**, 83-94.
- Li, J., Pancost, R.D., Naafs, B.D.A., Yang, H., Zhao, C., Xie, S., 2016. Distribution of glycerol dialkyl glycerol tetraether (GDGT) lipids in a hypersaline lake system. *Organic Geochemistry* **99**, 113-124.
- Lu, H., Liu, W., Wang, H., Wang, Z., 2016. Variation in 6-methyl branched glycerol dialkyl glycerol tetraethers in Lantian loess–paleosol sequence and effect on paleotemperature reconstruction. *Organic Geochemistry* **100**, 10-17.
- Manuilova, E., Schuetzenmeister, A., Model, F., 2014. Method Comparison Regression. CRAN, <https://cran.r-project.org/web/packages/mcr/index.html>.
- McMaster, G.S., Wilhelm, W.W., 1997. Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology* **87**, 291-300.
- Naafs, B.D.A., Inglis, G.N., Pancost, R.D., 2017. Introducing global peat-specific temperature and pH calibrations based on brGDGT bacterial lipids. *Geochimica et Cosmochimica Acta* **accepted pending minor revisions**.
- Pancost, R.D., Taylor, K.W.R., Inglis, G.N., Kennedy, E.M., Handley, L., Hollis, C.J., Crouch, E.M., et al., 2013. Early Paleogene evolution of terrestrial climate in the SW Pacific, Southern New Zealand. *Geochemistry, Geophysics, Geosystems* **14**, 5413-5429.
- Pearson, E.J., Juggins, S., Talbot, H.M., Weckström, J., Rosén, P., Ryves, D.B., Roberts, S.J., et al., 2011. A lacustrine GDGT-temperature calibration from the Scandinavian Arctic to Antarctic: Renewed potential for the application of GDGT-paleothermometry in lakes. *Geochimica et Cosmochimica Acta* **75**, 6225-6238.
- Peterse, F., van der Meer, M.T.J., Schouten, S., Jia, G., Ossebaard, J., Blokker, J., Sinninghe Damsté, J.S., 2009. Assessment of soil *n*-alkane δD and branched tetraether membrane lipid distributions as tools for paleoelevation reconstruction. *Biogeosciences* **6**, 2799-2807.
- Peterse, F., van der Meer, J., Schouten, S., Weijers, J.W.H., Fierer, N., Jackson, R.B., Kim, J.-H., et al., 2012. Revised calibration of the MBT–CBT paleotemperature proxy

based on branched tetraether membrane lipids in surface soils. *Geochimica et Cosmochimica Acta* **96**, 215-229.

Peterse, F., Martínez-García, A., Zhou, B., Beets, C.J., Prins, M.A., Zheng, H., Eglinton, T.I., 2014. Molecular records of continental air temperature and monsoon precipitation variability in East Asia spanning the past 130,000 years. *Quaternary Science Reviews* **83**, 76-82.

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA (USA).

Russell, N.J., 1984. Mechanisms of thermal adaptation in bacteria: blueprints for survival. *Trends in Biochemical Sciences* **9**, 108-112.

Schlitzer, R., 2015. Ocean Data View. <http://odv.awi.de>.

Schoon, P.L., de Kluijver, A., Middelburg, J.J., Downing, J.A., Sinninghe Damsté, J.S., Schouten, S., 2013. Influence of lake water pH and alkalinity on the distribution of core and intact polar branched glycerol dialkyl glycerol tetraethers (GDGTs) in lakes. *Organic Geochemistry* **60**, 72-82.

Schouten, S., Hopmans, E.C., Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Organic Geochemistry* **54**, 19-61.

Sinninghe Damsté, J.S., Hopmans, E.C., Pancost, R.D., Schouten, S., Geenevasen, J.A.J., 2000. Newly discovered non-isoprenoid glycerol dialkyl glycerol tetraether lipids in sediments. *Chemical Communications*, 1683-1684.

Sinninghe Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Weijers, J.W.H., Foesel, B.U., Overmann, J., Dedysh, S.N., 2011. 13,16-Dimethyl Octacosanedioic Acid (*iso*-Diabolic Acid), a Common Membrane-Spanning Lipid of *Acidobacteria* Subdivisions 1 and 3. *Applied and Environmental Microbiology* **77**, 4147-4154.

Weijers, J.W.H., Schouten, S., Hopmans, E.C., Geenevasen, J.A.J., David, O.R.P., Coleman, J.M., Pancost, R.D., et al., 2006a. Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environmental Microbiology* **8**, 648-657.

Weijers, J.W.H., Schouten, S., Spaargaren, O.C., Sinninghe Damsté, J.S., 2006b. Occurrence and distribution of tetraether membrane lipids in soils: Implications for the use of the TEX₈₆ proxy and the BIT index. *Organic Geochemistry* **37**, 1680-1693.

Weijers, J.W.H., Schefuß, E., Schouten, S., Damsté, J.S.S., 2007a. Coupled Thermal and Hydrological Evolution of Tropical Africa over the Last Deglaciation. *Science* **315**, 1701-1704.

Weijers, J.W.H., Schouten, S., van den Donker, J.C., Hopmans, E.C., Sinninghe Damsté, J.S., 2007b. Environmental controls on bacterial tetraether membrane lipid distribution in soils. *Geochimica et Cosmochimica Acta* **71**, 703-713.

- Weijers, J.W.H., Panoto, E., van Bleijswijk, J., Schouten, S., Rijpstra, W.I.C., Balk, M., Stams, A.J.M., et al., 2009. Constraints on the Biological Source(s) of the Orphan Branched Tetraether Membrane Lipids. *Geomicrobiology Journal* **26**, 402-414.
- Weijers, J.W.H., Bernhardt, B., Peterse, F., Werne, J.P., Dungait, J.A.J., Schouten, S., Sinninghe Damsté, J.S., 2011. Absence of seasonal patterns in MBT–CBT indices in mid-latitude soils. *Geochimica et Cosmochimica Acta* **75**, 3179-3190.
- Xiao, W., Xu, Y., Ding, S., Wang, Y., Zhang, X., Yang, H., Wang, G., et al., 2015. Global calibration of a novel, branched GDGT-based soil pH proxy. *Organic Geochemistry* **89–90**, 56-60.
- Yang, H., Lü, X., Ding, W., Lei, Y., Dang, X., Xie, S., 2015. The 6-methyl branched tetraethers significantly affect the performance of the methylation index (MBT') in soils from an altitudinal transect at Mount Shennongjia. *Organic Geochemistry* **82**, 42-53.
- Zell, C., Kim, J.H., Balsinha, M., Dorhout, D., Fernandes, C., Baas, M., Sinninghe Damsté, J.S., 2014. Transport of branched tetraether lipids from the Tagus River basin to the coastal ocean of the Portuguese margin: consequences for the interpretation of the MBT'/CBT paleothermometer. *Biogeosciences* **11**, 5637-5655.
- Zheng, Y., Li, Q., Wang, Z., Naafs, B.D.A., Yu, X., Pancost, R.D., 2015. Peatland GDGT records of Holocene climatic and biogeochemical responses to the Asian Monsoon. *Organic Geochemistry* **87**, 86-95.

Figure captions

Figure 1; World map with topography and location of soils used in this study, created using Ocean Data View (Schlitzer, 2015).

Figure 2; Fractional abundance of the brGDGT-Ia, -IIa, -IIIa, -IIa', and -IIIa' in the global soil sample data set versus mean annual air temperature (MAAT, top row), mean annual precipitation (MAP, middle row), and moisture index (MI, bottom row). Climatic parameters are obtained using PeatStash. Linear regressions are shown for those brGDGTs which relative abundance has a linear correlation coefficient (R^2) of at least 0.2. Zero values (below detection limit) are not included. Correlation of other brGDGTs to these parameters is not significant ($R^2 < 0.2$) and not shown.

Figure 3; Same as figure 2, but now for mean warmest month temperature (MWMT) and growing degree days above zero (GDD₀). Samples from soils with MAAT < 5 °C are highlighted in red in the MAAT plots.

631

632 Figure 4; MBT_{5me}' plotted versus a) mean annual air temperature (MAAT) and d)
633 growing degree days above 0 °C (GDD_0) together with Deming regression (purple
634 line) and simple linear regression (black dotted line). Also shown are the residuals for
635 both the Deming (b and e) and multiple linear regressions (c and f). Gray area in the
636 residual plots indicated missed variation because MBT_{5me}' reached 1.

637

638 Fig. 5; Histograms of pH values from mineral soils, showing that samples with IR_{6me}
639 < 0.5 are predominantly from acidic soils.

640

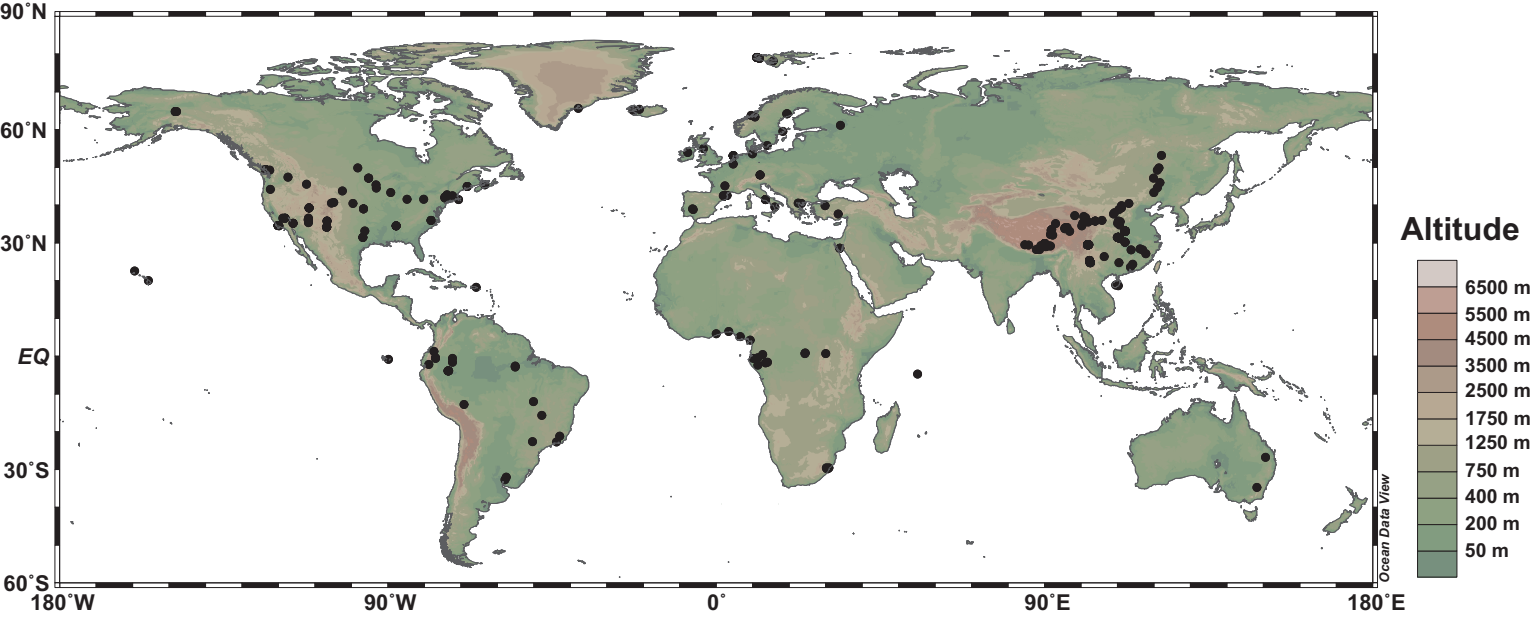
641 Fig. 6: The correlation coefficient (R^2) between MAAT and MBT_{5me}' versus the IR_{6me}
642 cut-off value as well as number of soils in each dataset. The total dataset (IR_{6me} cut-
643 off = 1) has a R^2 of 0.6 and consists of 350 samples.

644

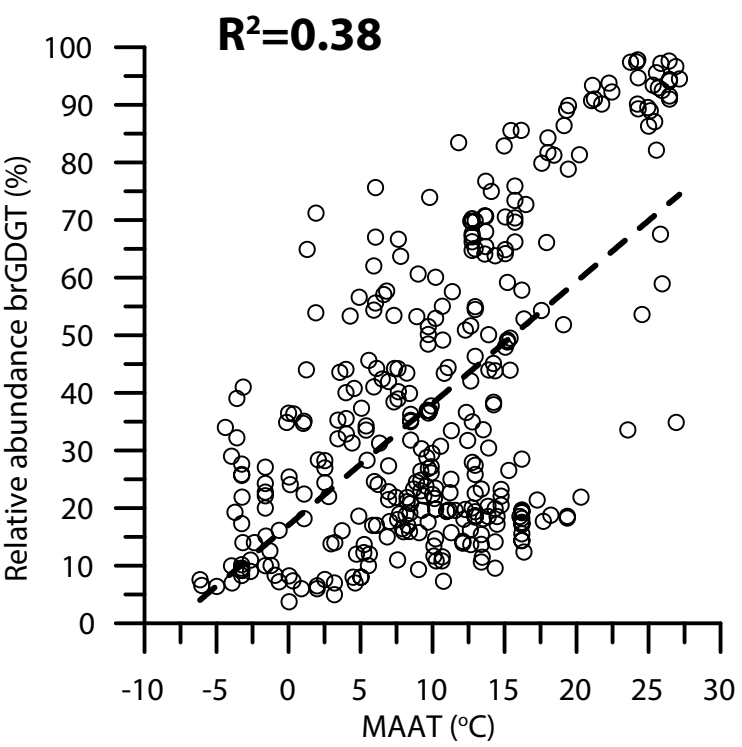
645 Fig. 7; Relative abundance of brGDGT-Ia, -IIa, and -IIIa versus MAAT and GDD_0 for
646 the complete soil data set (black, $n=350$) and soil samples with $IR_{6me} < 0.5$ (pink,
647 $n=177$).

648

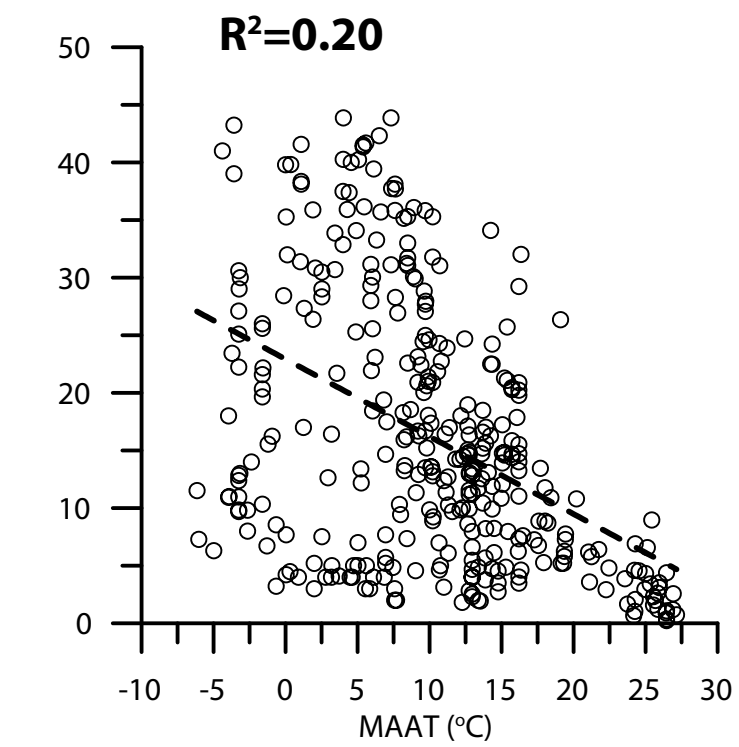
649 Figure 8; MBT_{5me}' of samples with $IR_{6me} < 0.5$ plotted versus a) mean annual air
650 temperature (MAAT) and d) growing degree days above 0 °C (GDD_0) together with
651 Deming regression (purple line) and simple linear regression (black dotted line). Also
652 shown are the residuals for both the Deming (b&e) and multiple linear regressions
653 (e&f). Gray area in the residual plots indicated missed variation because MBT_{5me}'
654 reached 1.



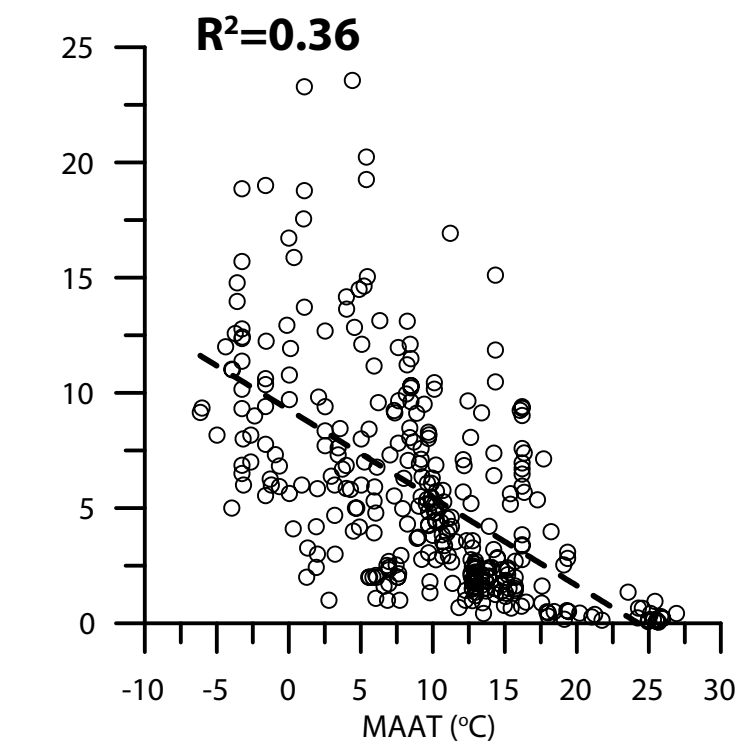
Ia



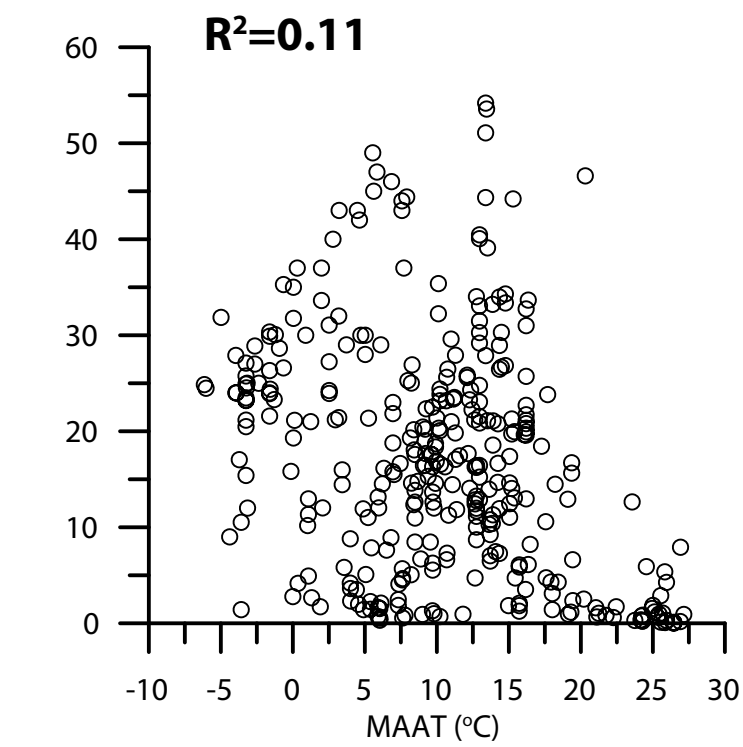
IIa



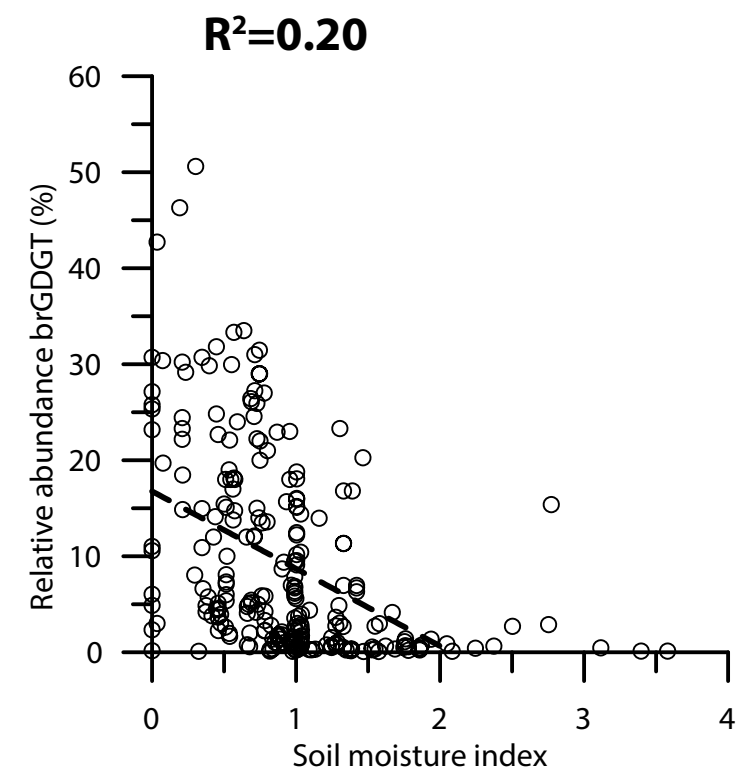
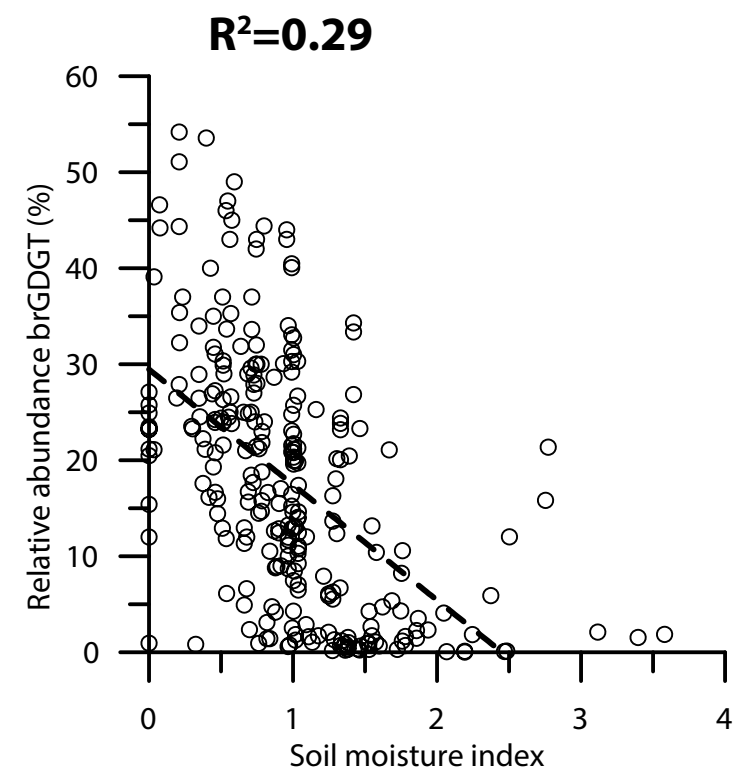
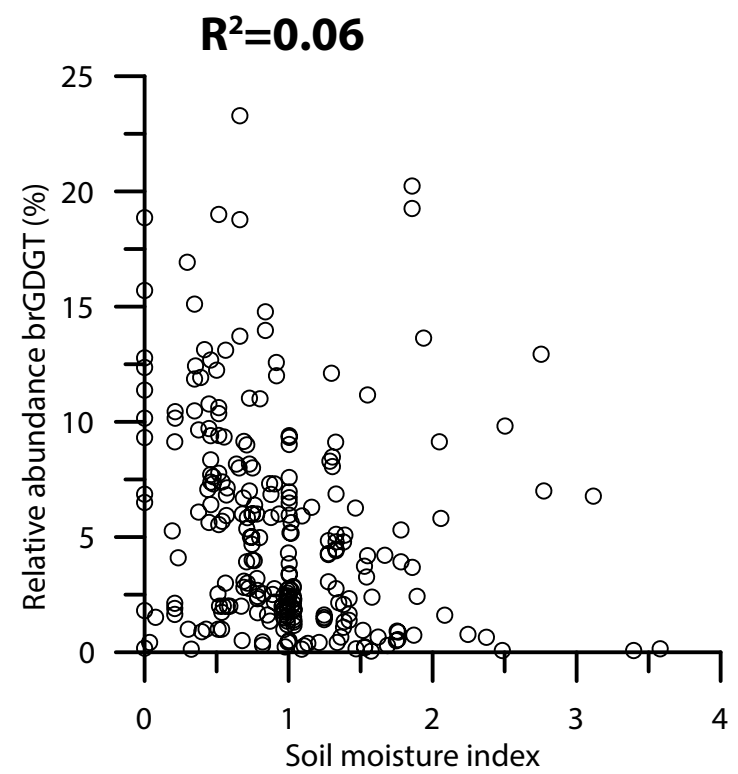
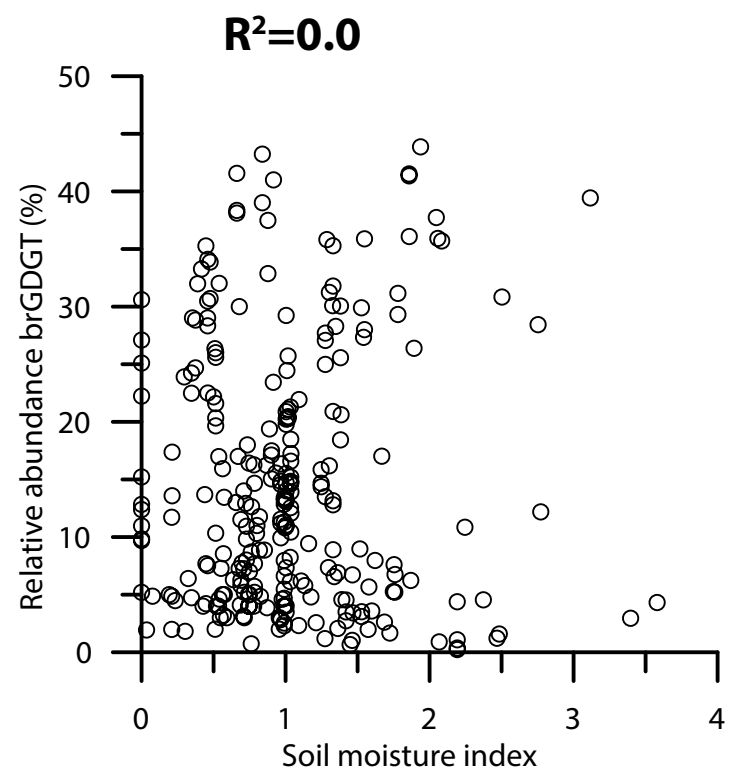
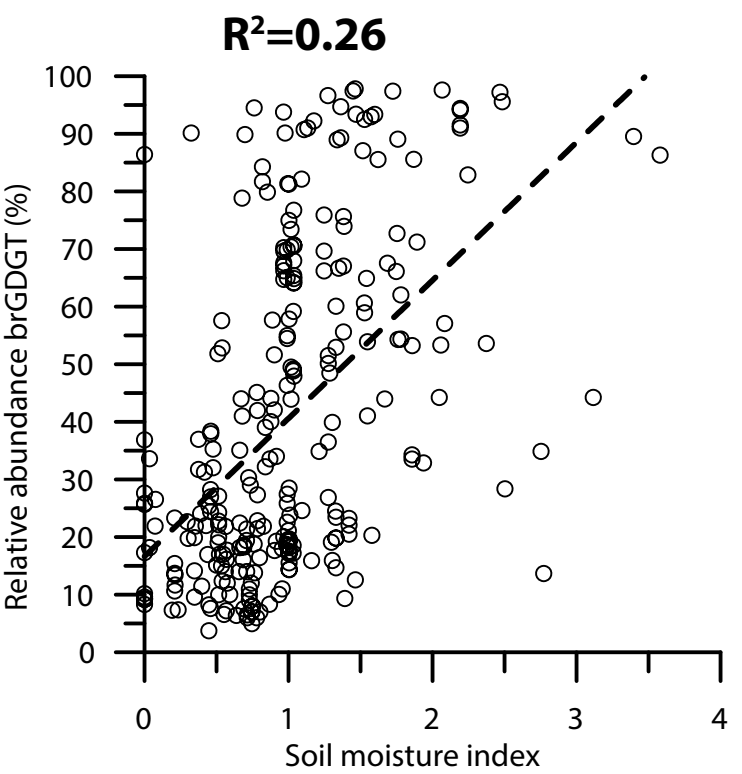
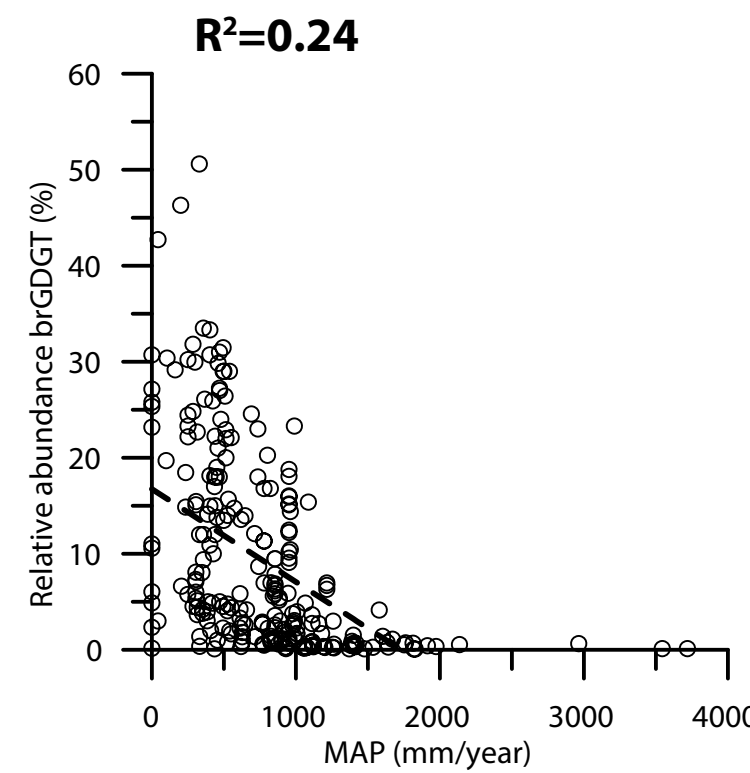
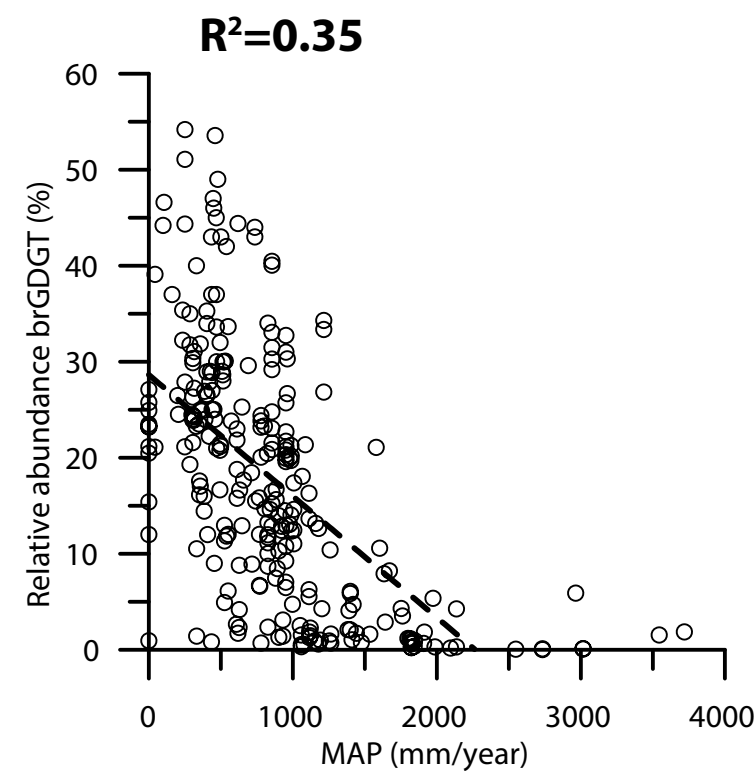
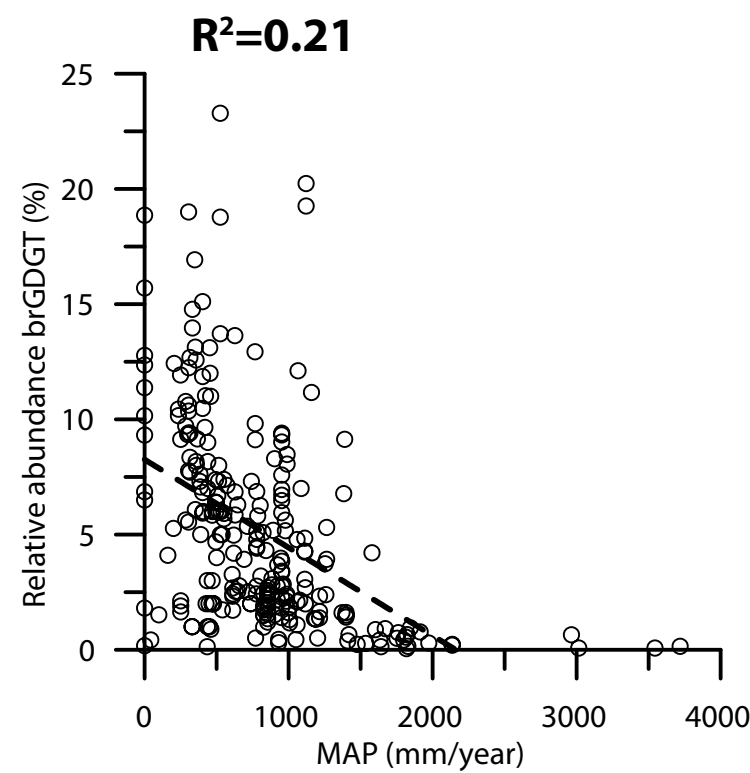
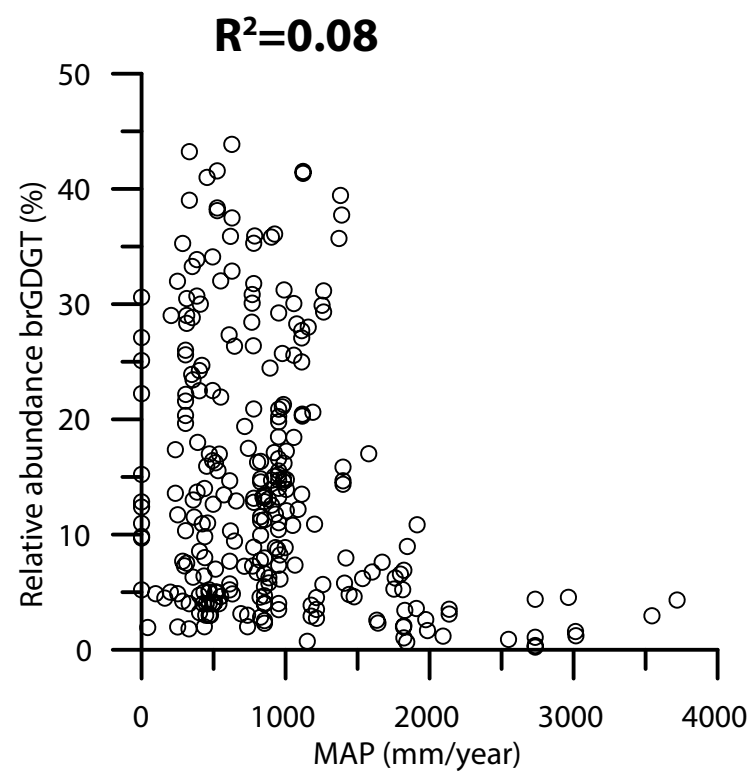
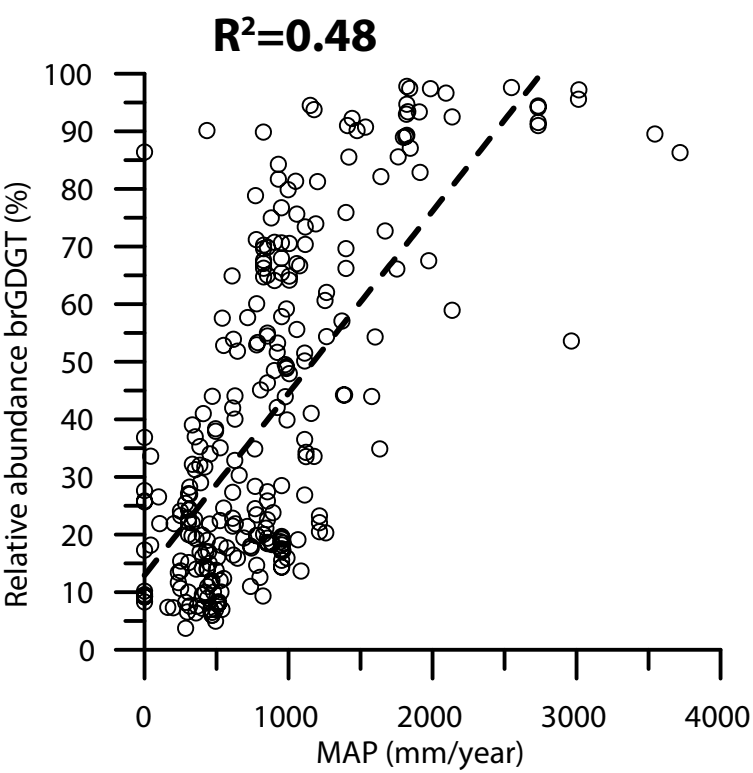
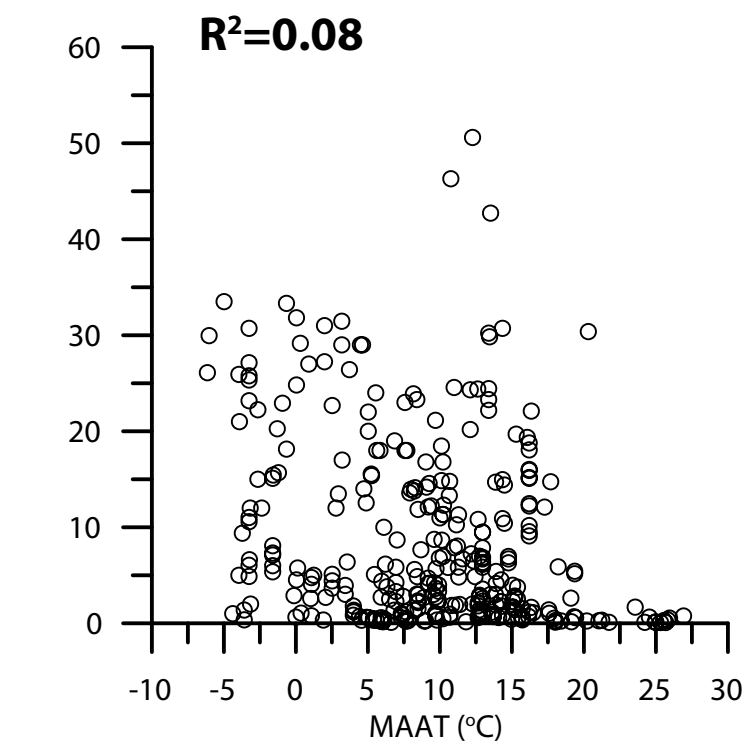
IIIa

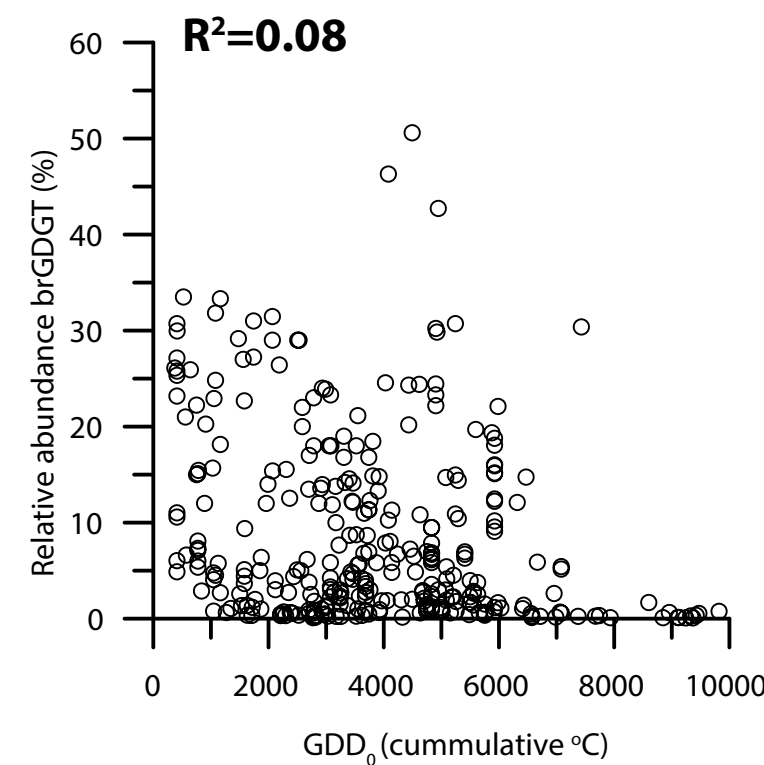
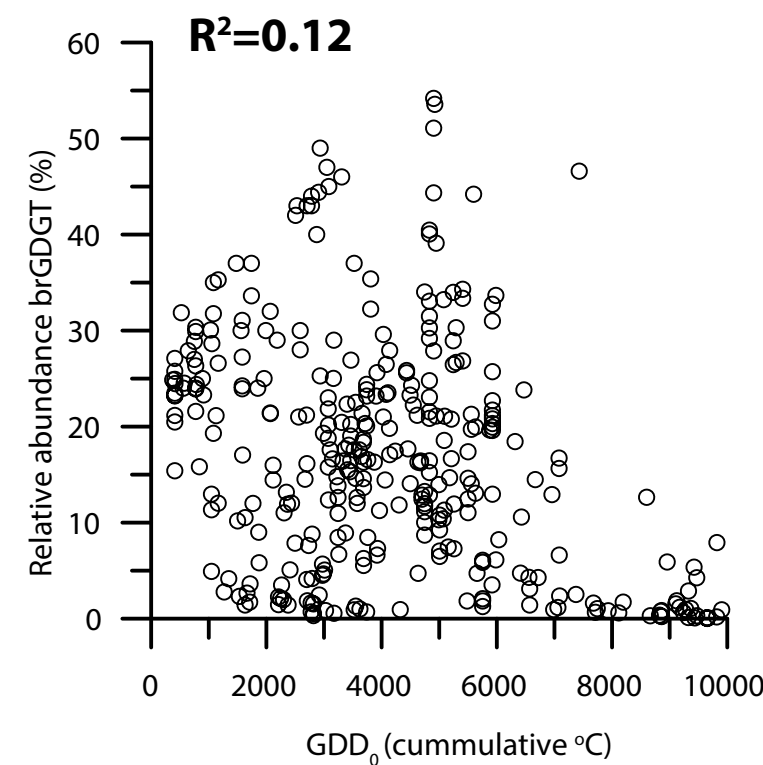
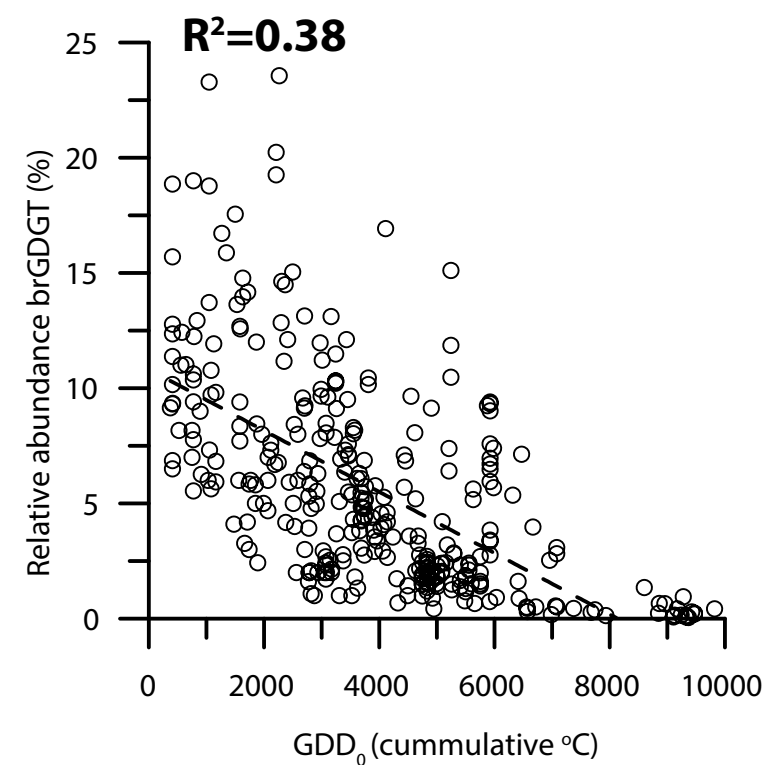
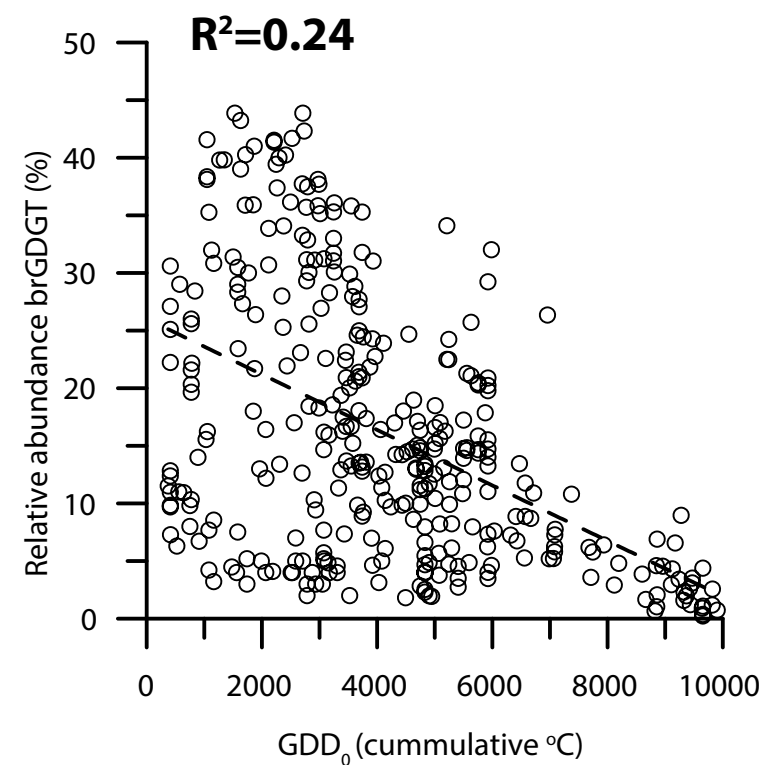
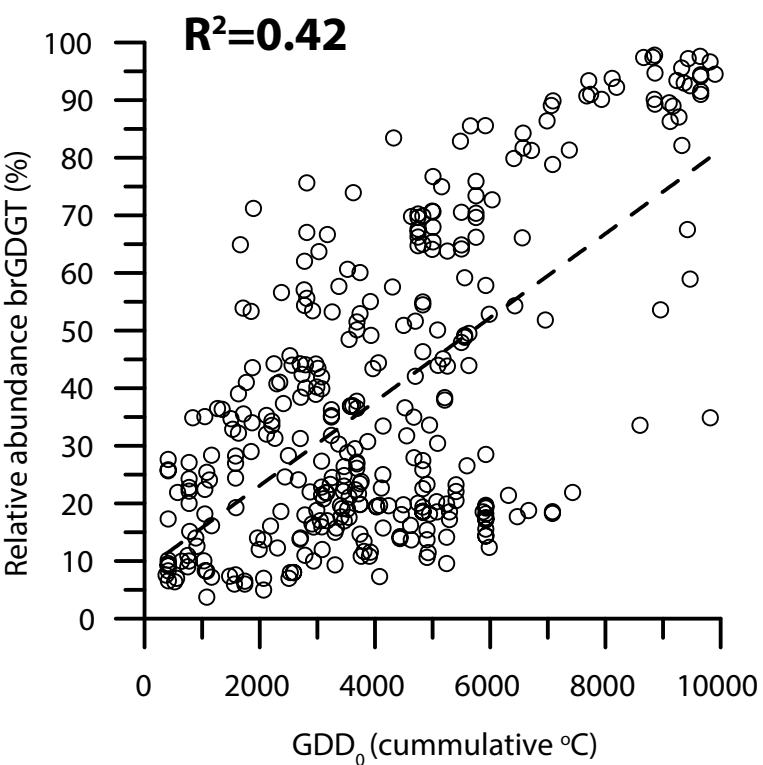
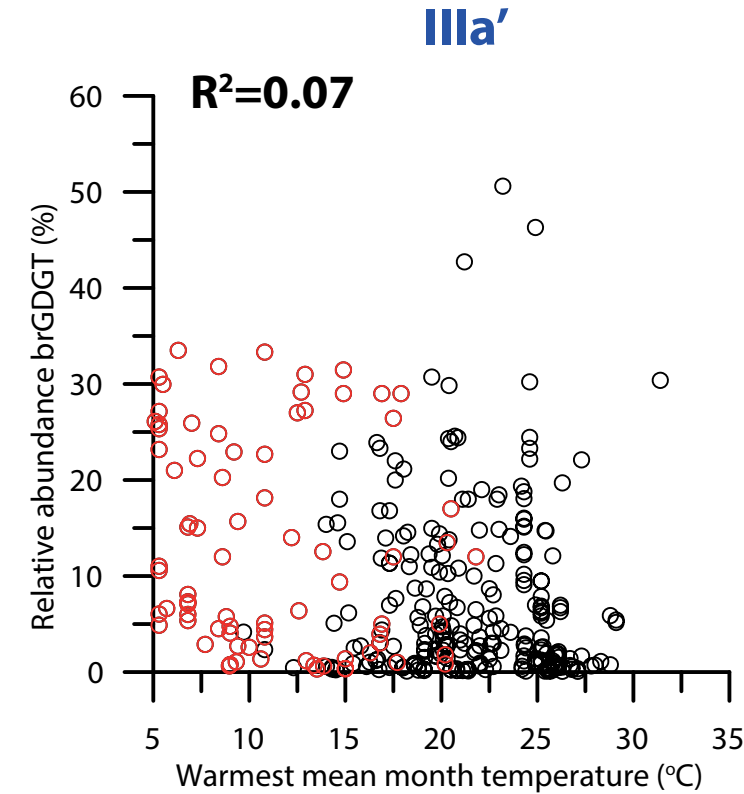
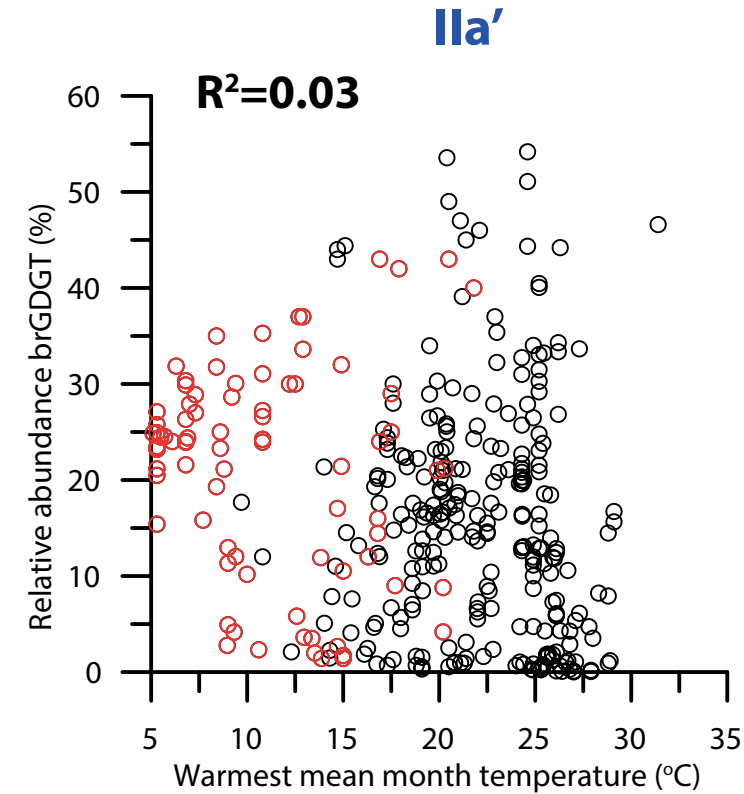
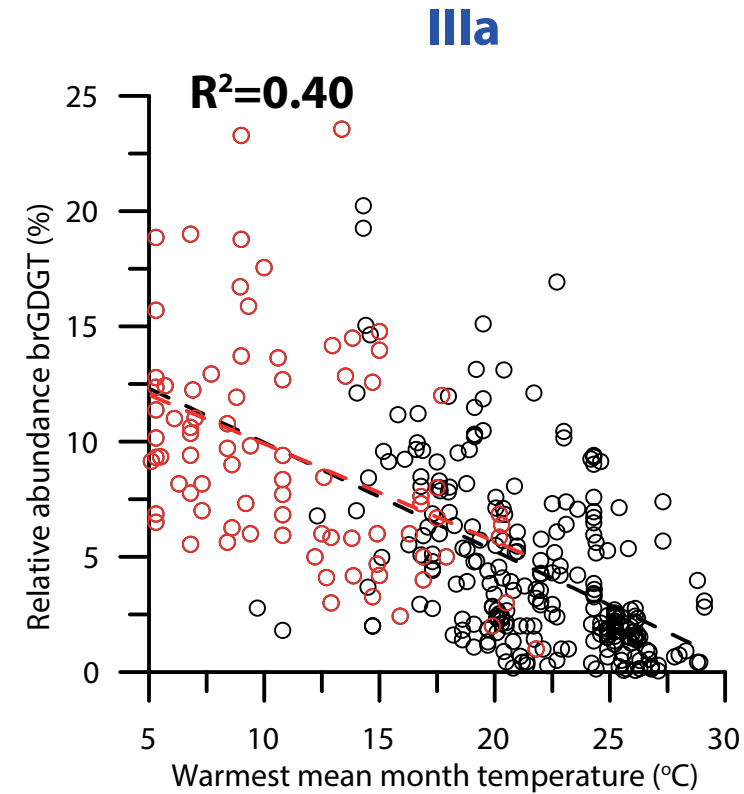
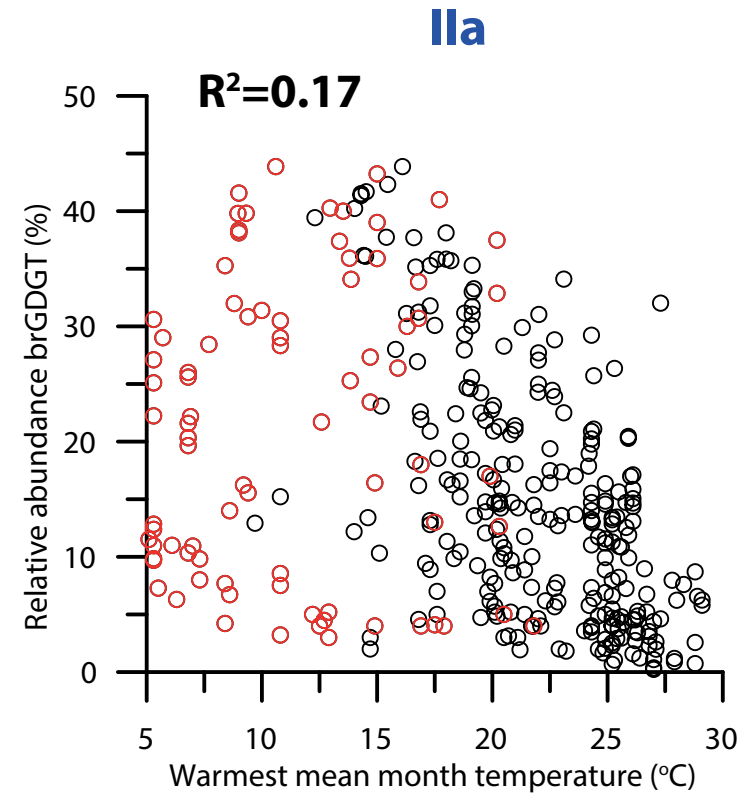
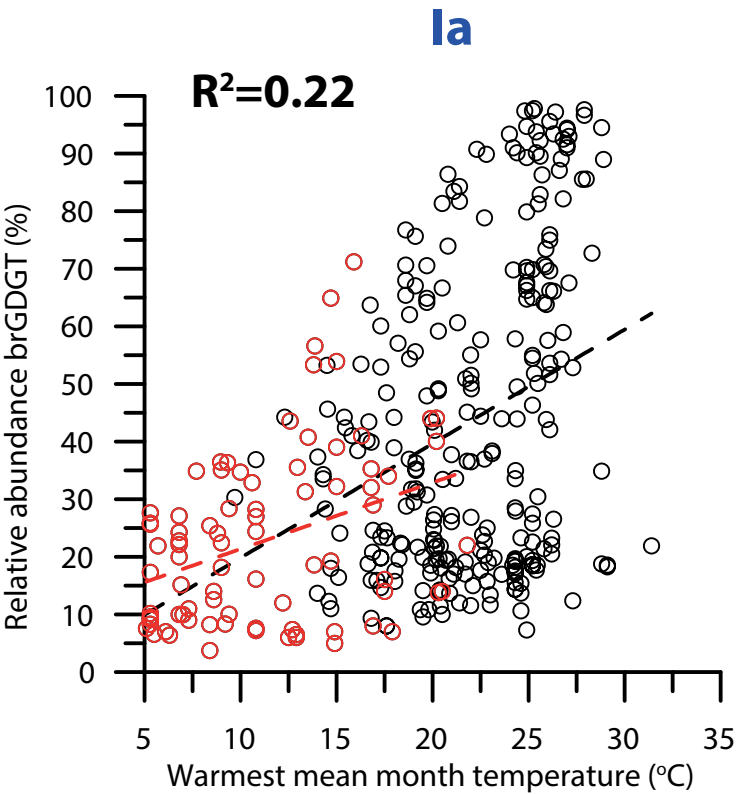


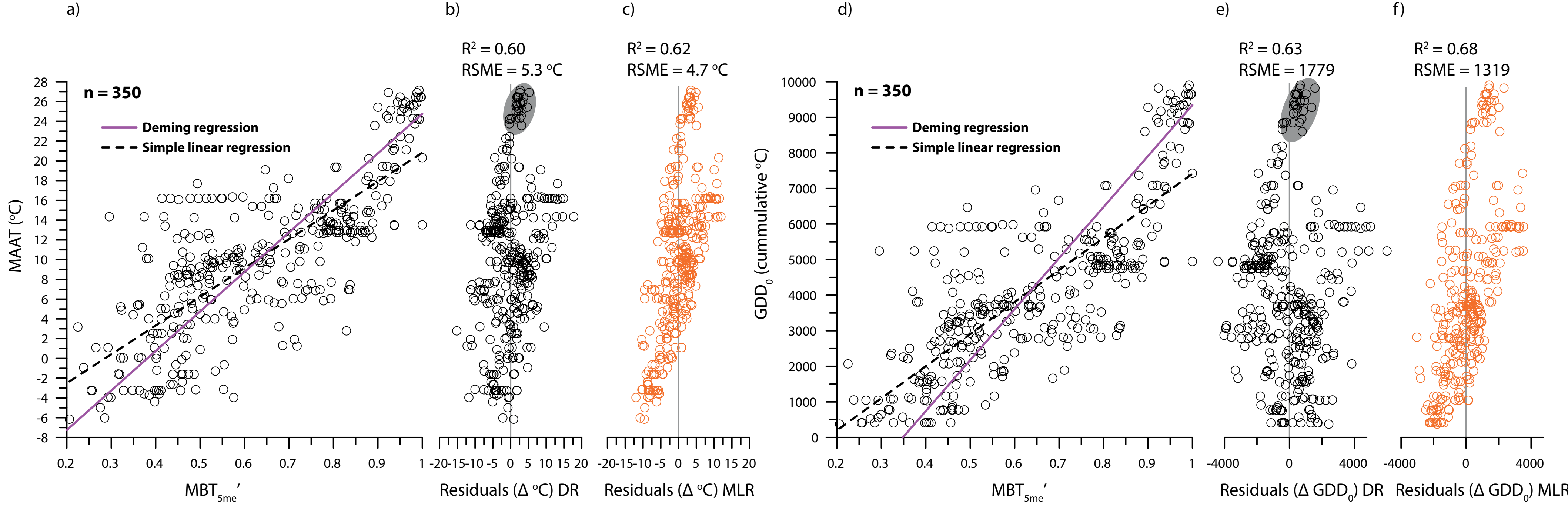
IIa'



IIIa'

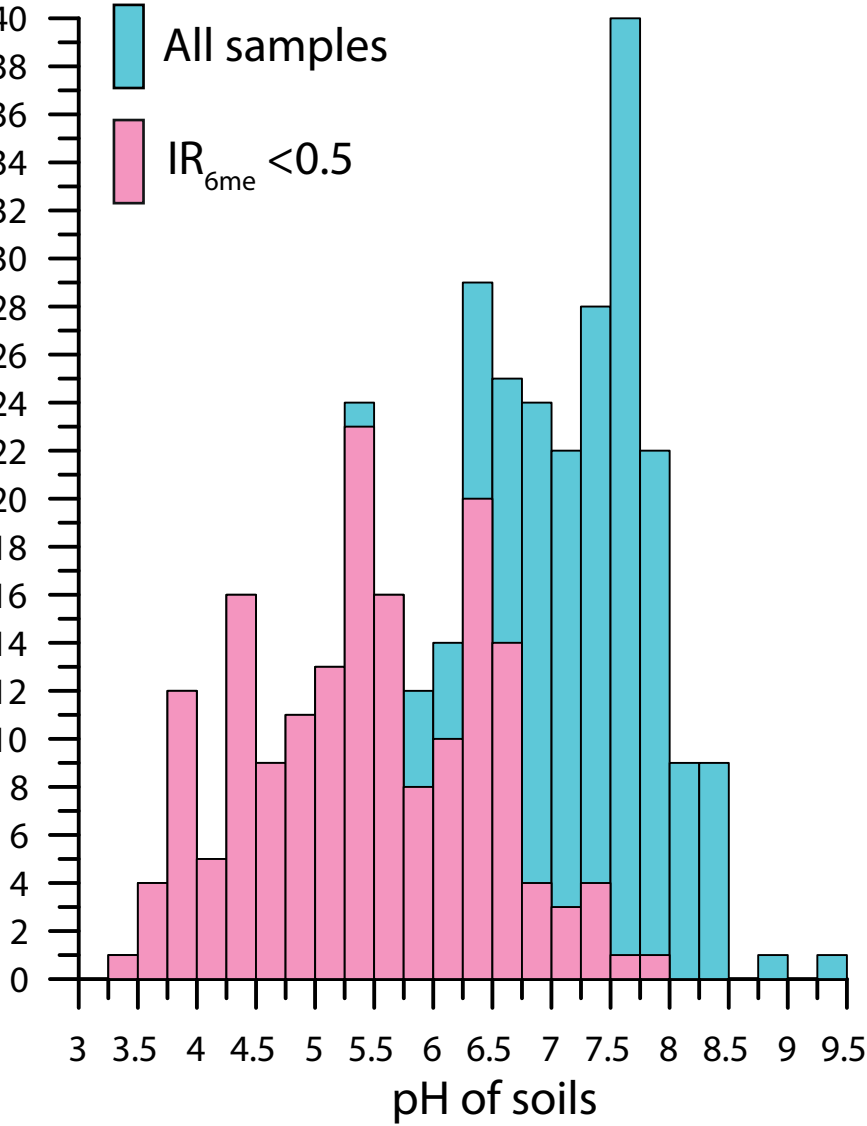


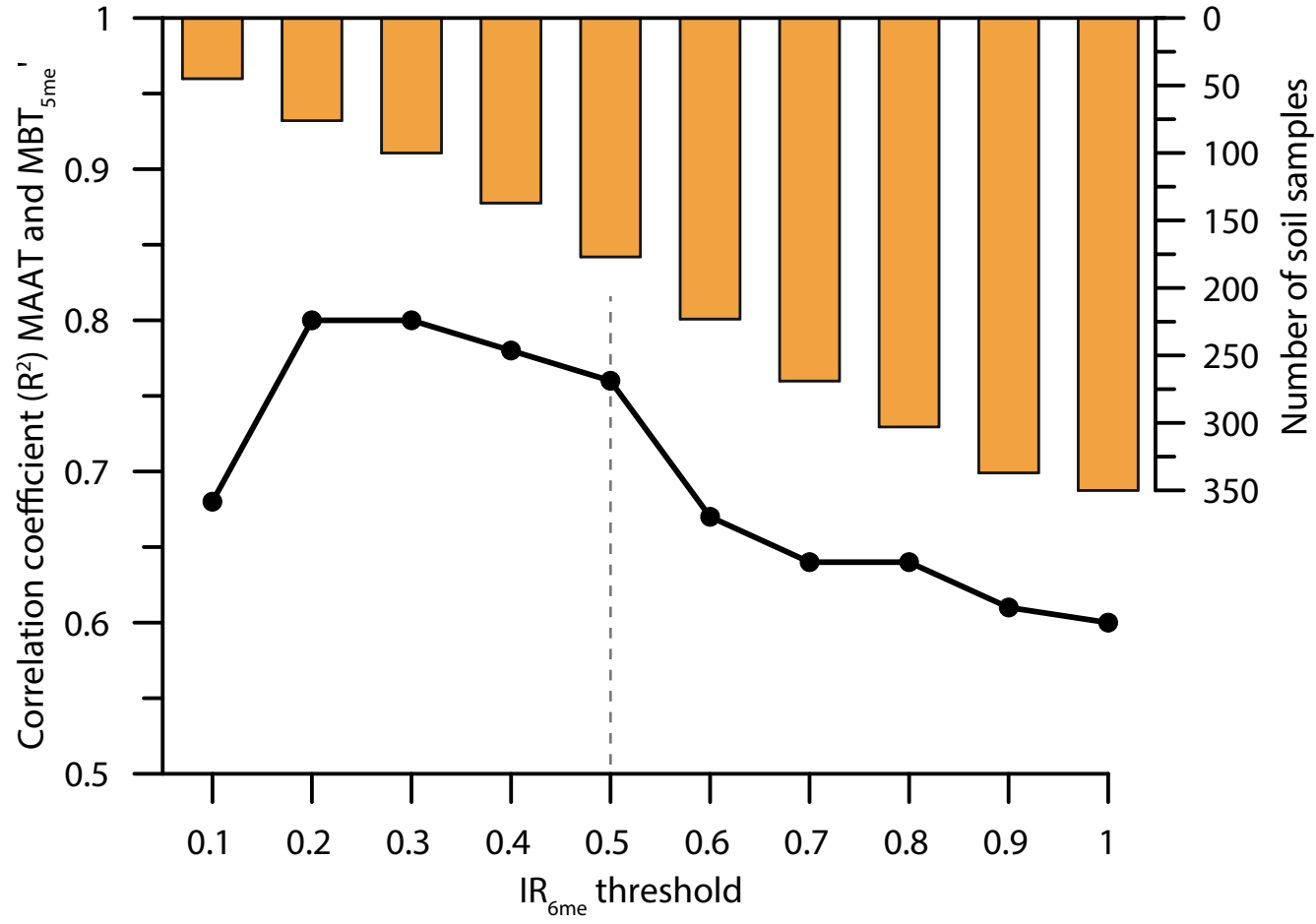


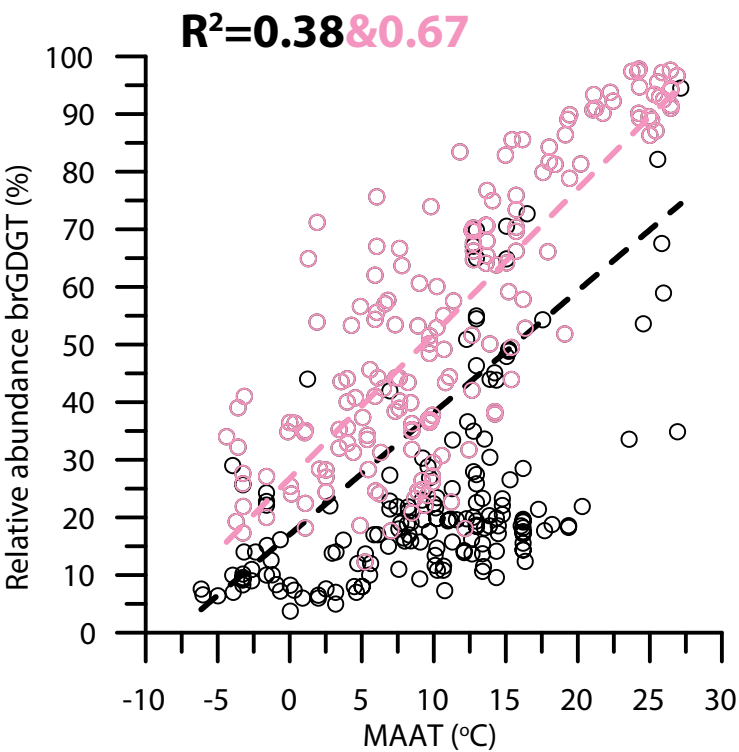
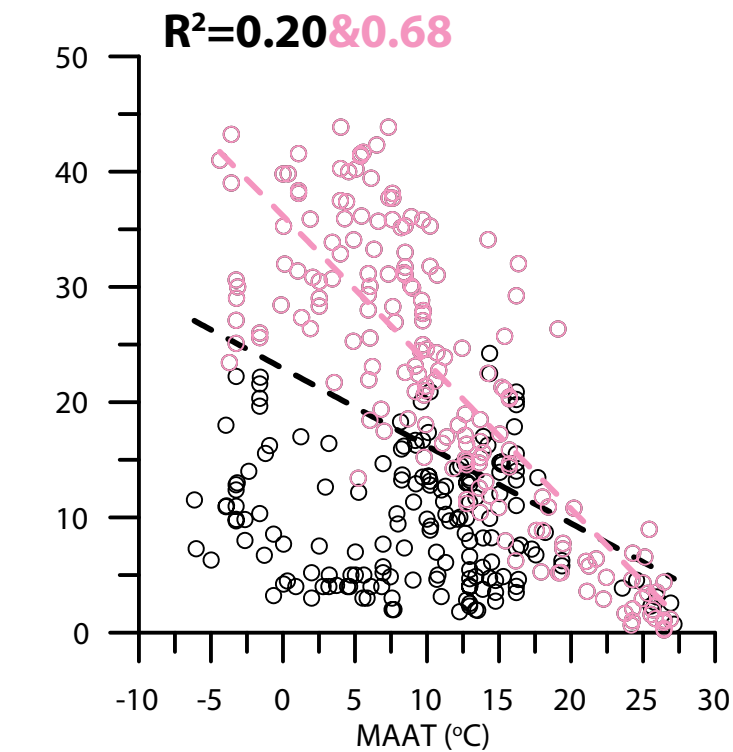


Number of soil samples

All samples
 $IR_{6me} < 0.5$





Ia**Ila****IIla**